



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

**NÁSTROJ PRO SPRÁVU DATABÁZE SEKVENČNÍCH
TYPŮ PATOGENNÍCH BAKTERIÍ**

MANAGEMENT TOOL FOR SEQUENCE TYPES DATABASE OF PATHOGENIC BACTERIA

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Zuzana Františová

VEDOUCÍ PRÁCE

SUPERVISOR

Mgr. Ing. Karel Sedlář

BRNO 2018

Bakalářská práce

bakalářský studijní obor **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Zuzana Františová

ID: 185948

Ročník: 3

Akademický rok: 2017/18

NÁZEV TÉMATU:

Nástroj pro správu databáze sekvenčních typů patogenních bakterií

POKYNY PRO VYPRACOVÁNÍ:

1) Zpracujte literární rešerši o základních metodách pro typizaci bakterií se zaměřením na metody Multi Locus Sequence Typing a Minim Typing. 2) Prostudujte možnosti a využití laboratorních technik, především sekvenace DNA a real-time PCR, pro rychlou typizaci bakterií. 3) Navrhněte a v prostředí Matlab implementujte základní převodní klíč pro překlad sekvenčního typu na příslušející melt typ umožňující srovnání sekvenčních a PCR dat. 4) Nástroj doplňte o možnost vytvářet a spravovat databáze melt typů, přičemž vytvořte i uživatelsky přívětivé grafické rozhraní. 5) Nástroj otestujte v laboratorním provozu ve fakultní nemocnici Brno a odstraňte případné nedostatky. 6) Proveďte diskusi a zhodnocení výsledků.

DOPORUČENÁ LITERATURA:

[1] ANDERSSON, Patiyan, et al. Minim Typing – A Rapid and Low Cost MLST Based Typing Tool for Klebsiella pneumoniae. PLoS ONE. 2012, 7(3), e33530.

[2] BRHELOVA, Eva, et al. Validation of Minim typing for fast and accurate discrimination of extended-spectrum, beta-lactamase-producing Klebsiella pneumoniae isolates in tertiary care hospital. Diagnostic Microbiology and Infectious Disease. 2016, 86(1), 44-49.

Termín zadání: 5.2.2018

Termín odevzdání: 25.5.2018

Vedoucí práce: Mgr. Ing. Karel Sedlář

Konzultant:

prof. Ing. Ivo Provazník, Ph.D.
předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Zveřejnění této práce je odloženo v souladu s ustanovením §47b odst. 4. zákona č. 111/1998 Sb. (Zákon o vysokých školách) v platném znění.

Důvodem odložení zveřejnění práce je ochrana právem chráněných zájmů Centra molekulární biologie a genové terapie FN Brno, IČ 652 69 705, které za účelem zpracování této závěrečné práce poskytlo autorovi své know-how a důvěrné informace vztahující se ke zpracovávané problematice.

Abstrakt

Předmětem této bakalářské práce je návrh a vytvoření nástroje pro správu databáze dosud známých sekvenčních typů patogenních bakterií. Základem práce je návrh převodního klíče nové techniky pro typizaci patogenních bakterií vytvořené ve spolupráci s Centrem molekulární biologie a genové terapie Interní hematologické a onkologické kliniky Fakultní nemocnice Brno. Metoda je založena na překladu sekvenčních typů na příslušející melt typ. V prvních dvou kapitolách jsou diskutovány výhody a nevýhody nejpoužívanějších molekulárně-biologických metod a z nich vycházejících metod pro bakteriální typizaci. Další tři kapitoly pak představují návrh převodního klíče pro novou metodu, jeho implementaci do prostředí MATLAB a výsledný nástroj složený ze dvou programů. Výpočet převodního klíče je součástí prvního z programů, druhý program umožňuje zpětně vyhledat konkrétní melt a příslušný sekvenční typ, a tak snadno srovnat sekvenční a PCR data.

Klíčová slova

sekvenční typ; převodní klíč; melt typ; typizace dle sekvencí mnoha lokusů; mini typizace

Abstract

The objective of this bachelor thesis is to design and create a management tool for sequence types database of pathogenic bacteria. The basis of this thesis is the designing a transfer key for the new molecular technique for typing of bacteria, which was created in cooperation with Centre of Molecular Biology and Gene Therapy, Department of Internal Medicine – Haematooncology, University Hospital Brno. The new method is based on translation sequence type to appropriate melt type. In the first two chapters, the advantages and disadvantages of the most used methods in molecular biology and the most favorite typing methods are discussed. Another three chapters introduce designing of the transfer key, its implementation in the MATLAB environment and the final management tool consisted of two programmes. The first program includes transfer key computation and the second program allows particular melt type and appropriate sequence types retrieval, which is an easy way to compare sequential and PCR data.

Keywords

sequence type; transfer key; melt type; Multi Locus Sequence Typing; Minim Typing

FRANTIŠOVÁ, Z. *Nástroj pro správu databáze sekvenčních typů patogenních bakterií*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 48 s. Vedoucí bakalářské práce Mgr. Ing. Karel Sedlář.

Prohlášení

„Prohlašuji, že svou bakalářskou práci na téma Nástroj pro správu databáze sekvenčních typů patogenních bakterií jsem vypracovala samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této semestrální práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne **25. května 2018**

.....
podpis autora(-ky)

Poděkování

Děkuji vedoucímu bakalářské práce Mgr. Ing. Karlu Sedláři za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé bakalářské práce.

V Brně dne **25. května 2018**

.....
podpis autora(-ky)

Obsah

Seznam obrázků	8
Seznam tabulek	8
Úvod.....	9
1 Základní molekulárně-biologické metody	10
1.1 Sekvence DNA	10
1.2 Elektroforetické techniky	13
1.3 Polymerázová řetězová reakce	14
2 Metody bakteriální typizace	16
2.1 Typizační metody založené na sekvenaci DNA	16
2.2 Typizační metody založené na elektroforéze	18
2.3 Typizační metody založené na PCR.....	19
3 Návrh převodního klíče	23
4 Implementace v prostředí MATLAB	27
5 Nástroj pro správu MS Access databáze	30
5.1 Propojení prostředí Matlab a Microsoft Access	31
5.2 Program pro správu MS Access databáze	32
5.3 Program pro vyhledávání.....	34
6 Zhodnocení.....	36
7 Závěr.....	39
Seznam použité literatury	41
Seznam použitých zkratk.....	46
Seznam elektronických příloh.....	48

Seznam obrázků

Obrázek 1.1 Sekvenace pomocí platformy SOLiD	11
Obrázek 1.2 Postup sekvenace metodou HeliScope	12
Obrázek 1.3 Přístroj pro SDS-polyakrylamidovou elektroforézu	13
Obrázek 1.4 HRM křivka	15
Obrázek 2.1 Postup MLST	17
Obrázek 2.2 Přehled veřejných databází PubMLST	18
Obrázek 2.3 Restrikční analýza	19
Obrázek 2.4 Fáze metody polymorfismu délky amplifikovaných fragmentů	20
Obrázek 2.5 PaLoc kódující dva geny pro bakteriální toxiny a tři regulační geny	21
Obrázek 3.1 Vývojový diagram načtení souborů a výpočtu vnitřní části převodního klíče	25
Obrázek 3.2 Vývojový diagram přiřazení MELT typu a uložení tabulky	26
Obrázek 4.1 Dialogové okno Načti FASTA soubory s geny	27
Obrázek 4.2 Ukázka vhodně popsaného a zarovnaného FASTA souboru	27
Obrázek 4.3 Ukázka části MLST databáze exportované do excel souboru	28
Obrázek 4.4 Dialogové okno Načti tabulku převodního klíče pro aktualizaci	28
Obrázek 4.5 Ukázka přesně dané struktury souboru s primery.....	29
Obrázek 4.6 Ukázka výsledného převodního klíče	29
Obrázek 5.1 Schéma nástroje	30
Obrázek 5.2 Uživatelské rozhraní	32
Obrázek 5.3 Zpětná vazba programu	32
Obrázek 5.4 Nový zápis primerů v Excel formátu	33
Obrázek 5.5 Schéma programu pro správu databáze	33
Obrázek 5.6 MS Access databáze	34
Obrázek 5.7 Uživatelské rozhraní programu pro vyhledávání.....	35
Obrázek 5.8 Nabídka možných kombinací alel	35
Obrázek 5.9 Podoba souboru vloženého při několikanásobném vyhledávání	35
Obrázek 6.1 Graf četností sekvenčních typů se společným melt typem	37

Seznam tabulek

Tabulka 1 Počty sekvenčních a melt typů pro jednotlivé bakterie.....	36
Tabulka 2 Absolutní (m_i) a relativní (r_i) četnost daného počtu ST se společným melt typem	37
Tabulka 3 Spearmanovy korelační koeficienty	38

Úvod

Studie po celém světě dokazují, že nozokomiální infekce (infekce z nemocničního prostředí) jsou významnou příčinou mortality. Mohou být způsobeny různými patogeny, kdy bakterie představují nejčastější příčinu. Četnost výskytu gastrointestinálních onemocnění způsobených patogenními bakteriemi stále roste [1]. V posledních letech je hojně diskutovaným tématem i růst rezistence jednotlivých kmenů vůči podávaným antibiotikům [2]. Tato fakta dokazují, že je nutné šíření patogenních bakterií sledovat a předcházet mu. Pro epidemiologické studie zabývající se touto problematikou je typizace patogenních bakterií základním nástrojem. Její hlavní úlohou je zjistit příbuznost mikrobiálních izolátů a stanovit, zda jsou epidemiologicky podobné izoláty geneticky příbuzné. Existuje celá řada různých technik, které je možné pro typizaci použít. Ty se často zásadně liší v ceně a ve způsobu provedení. Díky požadavkům vyšší rozlišovací schopnosti se již upouští od metod založených na fenotypu a do středu zájmu se dostávají především molekulárně-genetické metody.

První kapitola je věnována nejpoužívanějším molekulárně-biologickým metodám. Elektroforetické techniky, polymerázová řetězová reakce i sekvenace DNA nacházejí široké uplatnění ve výzkumu i praxi mnoha oblastí především chemie a molekulární biologie. V druhé kapitole jsou diskutovány výhody a nevýhody metod bakteriální typizace, které ze základních molekulárně-biologických metod vychází. Z finančních důvodů se však většinou typizace bakterií využívá spíše pro výzkum než pro rutinní práci v laboratorním provozu. Za tímto účelem se tedy neustále hledají nová řešení. Přitom je důležité, aby při snížení nákladů byla nová technika neméně spolehlivá než techniky stávající.

V dalších dvou kapitolách je pak popsán návrh převodního klíče pro novou metodu a jeho implementace do prostředí MATLAB. Základní převodní klíč byl navržen ve spolupráci s Centrem molekulární biologie a genové terapie Interní hematologické a onkologické kliniky Fakultní nemocnice Brno. Principem metody je překlad sekvenčního typu na příslušející melt typ. Samotný nástroj je složen ze dvou programů, jejichž popisu je věnována další kapitola. V kapitole je také diskutována volba MS Access databáze a úskalí při navazování spojení s prostředím Matlab. Výpočet převodního klíče je součástí prvního z programů. Algoritmus na základě načtených dat přiřadí sekvenčním typům příslušný melt typ a umožní uložení klíče do databáze. Při načtení nových dat lze klíč i aktualizovat. Druhý program je určen pro praktické využití databáze. Uživatel zpětně vyhledává na základě zvolených alel konkrétní melt a příslušný sekvenční typ. Tento postup umožňuje srovnání sekvenačních a PCR dat. Díky tomu je možné identifikovat patogenní bakterie rychle, efektivně a levně.

1 Základní molekulárně-biologické metody

Metody molekulární biologie umožňují analýzu biologicky významných molekul, zejména nukleových kyselin a bílkovin nesoucích genetickou informaci. Jejich rozvoj v posledních dvaceti letech umožnil detailní analýzu genetické informace patogenů i pacientů. Polymerázová řetězová reakce [3], elektroforetické techniky [4] a sekvenace DNA jsou dnes metody nacházející široké uplatnění ve výzkumu i praxi mnoha oblastí především chemie a molekulární biologie. Na nich jsou mimo jiné založeny i metody typizace.

1.1 Sekvenace DNA

První generace

První dvě metody sekvenace vznikly v druhé polovině 80. let. Maxam-Gilbertovo sekvenování je založeno na chemické modifikaci DNA a následném štěpení řetězce v místech modifikace nukleotidů [5]. Sangerovo sekvenování využívá při syntéze nového řetězce specifických vlastností DNA polymeráz [6]. V obou případech se tedy jedná o enzymové sekvenování. Přestože Maxam-Gilbertova metoda byla zpočátku oblíbenější než Sangerova díky schopnosti aplikace přímo na purifikovanou DNA bez nutnosti vytvoření jednovláknové DNA, dnes je již plně nahrazena další generací. Sangerova metoda však stále nachází uplatnění i v dnešní době při sekvenování menšího rozsahu.

Druhá generace

Sekvenátory druhé generace (častěji používané anglické označení Next Generation Sequencing, zkráceně NGS) dokázaly významně snížit dobu potřebnou k přečtení DNA díky masivnímu paralelnímu sekvenování až tisíců molekul DNA současně. Technologie jsou založené na PCR amplifikaci templátu. Tři nejznámější platformy Roche 454, Illumina a SOLiD byly uvedeny na trh začátkem 21. století [7]. Všechny mají společné tři základní kroky. Každá platforma však pracuje s jinými kombinacemi přístupů k těmto krokům. Platforma Roche 454 je z historického hlediska nejstarší platforma druhé generace, která již dnes není podporovaná. Princip založený na pyrosekvenování lze vysvětlit jako proces enzymatických reakcí, kdy se do nově vznikajícího řetězce začleňují jednotlivé báze a dochází k emisi světla z viditelné oblasti spektra. Kolik světla se uvolní, závisí na počtu nukleotidů, které se začlení. Proces se cyklicky opakuje a sekvence je odečítána z tzv. pyrogramu. Cena reagentů pro tuto technologii byla příliš vysoká. Největším negativem však byl problém s detekcí homopolymerních úseků DNA. V těchto úsecích byla metoda náchylná k větším chybám v určení počtu začleněných nukleotidů a z toho plynoucí inserce nebo delece spíše než substituce nukleotidu [8].

Platforma Illumina je v současnosti nejrozšířenější. Byla uvedena na trh v roce 2007. Pracuje na principu sekvenace syntézou ve spojení s tzv. můstkovou amplifikací. Při této technice dochází k hybridizaci templátové DNA na pevný povrch transparentní komůrky s využitím chemicky reverzibilně modifikovaných nukleotidů [8]. Výhodou Illumina je velmi

Platforma SOLiD je nejméně rozšířená. Technologie se od Illumina a Roche 454 liší tím, že je založena na sekvenování ligací (odtud i název platformy z anglického „Sequencing by Oligo Ligation and Detection“), které využívá hybridizaci krátkých fluorescenčně značených sond. Nejběžnějším typem chyby je opět substituce. Díky dvojitému čtení je také pomalejší než ostatní metody. Pro představu, přečtení dat o velikosti 3-4 Gbp zabere zhruba pět dní. Technika sekvenování je zobrazena na obrázku 1.1.

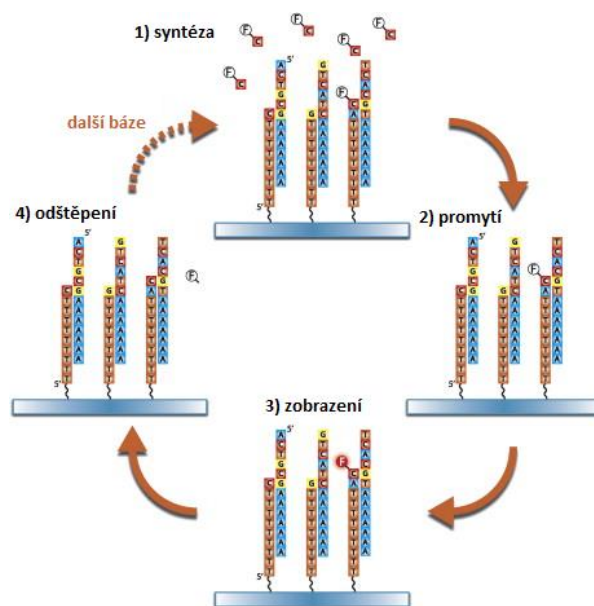


Třetí generace

11

redukce ceny, snížení chybovosti pramenící z amplifikace, možnost vyšší flexibility v délce čtení a přesnost kvantifikace DNA molekul, protože signál je zaznamenáván v reálném čase [8]. Pro tuto technologii je vyžadován velmi citlivý systém detekce světla, který musí být schopen identifikovat světlo z jediného fotonu [12]. Mezi platformy umožňující tuto techniku patří Helicos, PacBio SMRT a Oxford Nanopore.

Platforma Helicos již dnes není podporovaná a nikdy nebyla využita pro významné projekty, přesto její technologie stojí za zmínku. Pracuje na principu tzv. přesného sekvenování jednotlivých molekul DNA (z anglicky True Single Molecule Sequencing, zkráceně tSMS) a dokáže analyzovat mnoho milionů jednotlivých fragmentů DNA současně. Díky tomu je umožněna sekvenace rychlostí až 1 Gb za den. Před samotnou sekvenací probíhá štěpení DNA vzorku na úseky délky 100-200 bp a připojení adaptérů s fluorescenční značkou na polyadeninovém 3'-konci. Denaturované řetězce jsou na 3'-koncích hybridizovány příslušnými oligonukleotidy a následně uchycovány na povrchu destičky. Destička je schopna zachytit až 10^8 templátů na čtverečním centimetru. Pro sekvenaci je na destičku přidána DNA polymeráza a jeden ze čtyř fluorescenčně značených deoxynukleozidtrifosfátů. Polohové souřadnice zachycených řetězců jsou zaznamenány CCD kamerou a před dalším sekvenováním je fluorescenční značení odděleno a odplaveno [13]. Postup je znázorněn na obrázku 1.2.



Obrázek 1.2 Postup sekvenace metodou HeliScope (upraveno z [14])

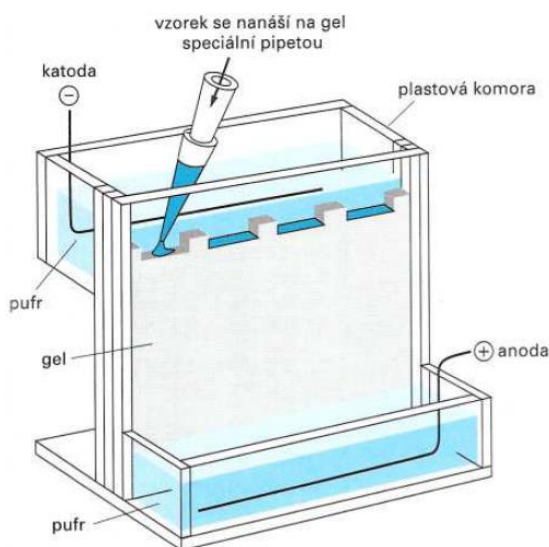
Pacific Biosciences SMRT je zkratka z anglického Single Molecule Real-Time Sequencing, což v překladu znamená sekvenace jednotlivých molekul v reálném čase. Klíčové jsou maličké nanojamky o velikosti asi 70 nm v průměru a 100 nm v hloubce, v nichž je vložena jediná molekula polymerázy, která je pozorována při syntéze vlákna DNA. Platforma přímo využívá rychlosti a efektivity této reakce a nevzniká nutnost vytváření shluků molekul DNA. Díky tomu, že sekvenátor nepozastavuje proces přirozené reakce polymerázy, je sekvenování velice rychlé a délka čtení velká. Navíc jamek mohou být až desetitisíce vedle sebe, čímž je

umožněno masivní paralelní sekvenování. Příprava vzorku je jednoduchá a rychlá, nicméně spojena s nákladným vybavením a činidly a je závislá na velkém prostoru laboratoře. Nevýhodou je poměrně velká chybovost sekvenování (nad 10 %) [15].

Oxford Nanopore je platforma umožňující sekvenace proteinovým nanopórem v membráně ze syntetického polymeru s velkým elektrickým odporem. Když je DNA vložena na povrch nanopóru s exonukleázou a na lipidové dvojvrstvě je změněno napětí, exonukleáza je schopna uštěhnout individuální nukleotidy z této DNA. Tyto individuální nukleotidy specificky narušují iontový proud probíhající pórem, čímž je možné je detekovat. Díky tomu, že k detekci jsou místo optické detekce využity změny elektrického prostředí, je možno dosáhnout velice nízké ceny za detekovanou bázi [15].

1.2 Elektroforetické techniky

Elektroforéza využívá schopnosti nabitých částic pohybovat se v elektrickém poli. Rychlost pohybu částic je závislá na velikosti náboje a velikosti molekuly. Na tomto principu dokážou molekuly o rozdílném náboji nebo hmotnosti izolovat. Jednotlivé techniky se pak liší podle použitého prostředí. Může se jednat o vodný roztok či hydrofilní porézní nosiče (např. neklížený papír, acetát celulózy, agarózový gel, polyakrylamidový gel, celulózu). Pro analýzu DNA, proteinu nebo enzymů se nejčastěji využívá elektroforéza v polyakrylamidovém nebo agarózovém gelu. Gelová elektroforéza může být buď nativní nebo tzv. SDS-gelová elektroforéza, při které dochází k denaturaci proteinů dodecylsíránem sodným (odtud zkratka SDS). Přístroj pro tuto techniku je znázorněn na obrázku 1.3 [4]. Při nativní gelové elektroforéze póry v gelu slouží jako molekulové síto. Velikost pórů je zvolena podle velikosti proteinů a nukleových kyselin. Speciální technikou je kapilární elektroforéza založená na principu pohybu elektricky nabitých látek rozpuštěných ve vodivém médiu a principu elektroosmózy umožňující separaci látek. K tomuto účelu se využívá kapilára naplněná elektrolytem nebo gelem [16].



Obrázek 1.3 Přístroj pro SDS-polyakrylamidovou elektroforézu (upraveno z [4])

1.3 Polymerázová řetězová reakce

Polymerázová řetězová reakce (anglicky Polymerase Chain Reaction, zkráceně PCR) je rychlá, jednoduchá a flexibilní molekulární technika umožňující vysoce selektivní namnožení zvolené sekvence DNA z templátové DNA pomocí jednoduché enzymatické reakce opakující se až milionkrát během několika hodin. Princip spočívá v tom, že DNA polymeráza je schopna navázat se na jednořetězcovou templátovou DNA jen s pomocí krátkého primeru (oligonukleotidu) a katalyzovat replikaci in vitro u sekvence nukleové kyseliny, kterou požadujeme [3].

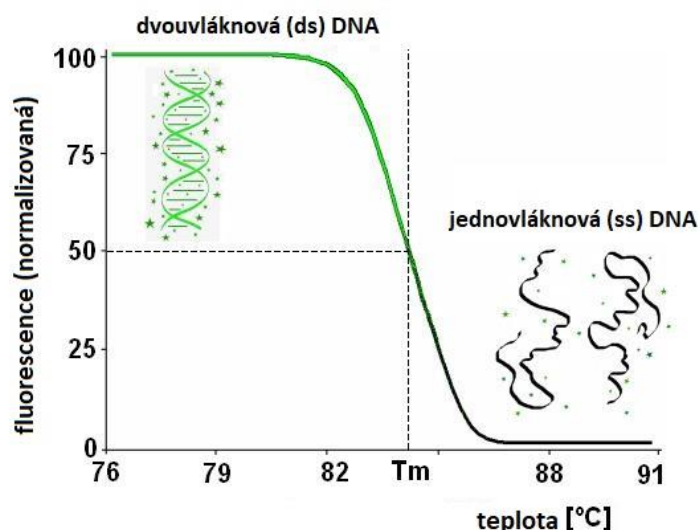
Dvouvláknová DNA je denaturována na jednotlivá vlákna pomocí vysoké teploty. Následně je k vybraným úsekům komplementárních vláken DNA připojen vždy jeden ze dvou druhů primerů. V tomto místě pak nasedá DNA polymeráza syntetizující nové komplementární vlákno. Díky tomu, že jsou DNA polymerázy termostabilní (odolávající vysokým teplotám), můžeme provádět syntézu DNA opakovaně v řadě cyklů, kdy jeden cyklus je složen ze tří kroků s odlišnými nároky na teplotu (denaturace dvou řetězcové DNA při 94-95 °C, připojení primerů k řetězcům DNA při 50 - 65 °C a syntéza nových řetězců DNA pomocí DNA polymerázy při 65 - 75 °C). Nevýhodou je extrémní citlivost metody, neboť hrozí riziko kontaminace z laboratorního prostředí (od bakterií, virů až po cizorodou DNA). Pro odhalení případné kontaminace a kontrolu amplifikace je nezbytné provádět negativní a pozitivní kontrolu. Problémem mohou být falešně negativní výsledky způsobené přítomností extracelulárních a intracelulárních PCR inhibitorů. Z tohoto důvodu byla velká pozornost soustředěna na vývoj rychlých a robustních metod izolace DNA [17]. Často využívané jsou také nejrůznější modifikace klasické PCR. Mezi ty nejznámější patří kvantitativní PCR (zkráceně qPCR), mnohonásobná (též multiplexní) PCR, inverzní PCR (zkráceně iPCR) a interrepetitivní PCR (zkráceně repPCR). Například principem interrepetitivní PCR je amplifikace repetitivních DNA sekvencí, které jsou přítomny v bakteriálním genomu. Ty se vyskytují v chromozomu mnoha bakterií v různých počtech a na různých pozicích. Pokud jsou nalezeny dva repetitivní elementy dostatečně blízko u sebe, mohou být využity k amplifikaci úseku DNA mezi nimi. Tyto oblasti se pak nazývají interrepetitivní fragmenty.

Polymerázová řetězová reakce v reálném čase

Varianta PCR umožňující přímou detekci a kvantifikaci PCR produktů v reálném čase (anglicky real-time PCR) je založena na principu neustálého snímání intenzity fluorescenčního signálu a okamžité analýzy díky přístroji, jenž umožňuje zároveň PCR [18]. Tato technika je tedy jednodušší a rychlejší, neboť amplikony se nemusí detekovat elektroforeticky. Další předností je také přesnost a citlivost [19]. Oproti klasické PCR je výsledek kvantifikace počátečního množství templátu na konečné množství PCR produktu mnohem spolehlivější díky zaznamenávání vzniku amplikonů v průběhu celé reakce. K detekci vznikajícího PCR produktu lze využít fluorescenční barviva, fluorescenčně značené hybridizační sondy nebo fluorescenčně značené primery [20]. Technika fluorescenčních barviv je nejjednodušší. Barviva se vloží mezi

báze DNA a narůstající fluorescence pak odpovídá vzrůstajícímu množství DNA ve vzorku. Technika značení primerů a hybridizačních sond využívá specifické molekuly fluoroforů emitujících světlo určité vlnové délky po předchozí absorpci světla odlišné vlnové délky. Vyzářená vlnová délka světla je vždy vyšší než absorbovaná. Dvojitě fluorescenčně značené sondy obsahují kromě fluoroforu také tzv. zhášec (molekulu přijímající energii z fluoroforu ve formě světla), který způsobuje rozptýlení buď ve formě tepla nebo světla s vyšší vlnovou délkou. Optimálního zhášení je dosaženo, když se absorpční spektrum zhášeče a emisní spektrum fluoroforu překrývají.

Analýza s vysokou rozlišovací schopností (z anglického High Resolution Melting, zkráceně HRM) je snadná post PCR metoda sloužící mimo jiné k detekci mutací. Metoda vyžaduje pouze standardní PCR reagentie a fluorescenční barvivo, které je vloženo do dvouvláknové DNA. Tímto krokem lze detekovat heteroduplexy, které vznikly po PCR reakci. Změny profilu křivek tání ve srovnání s negativní kontrolou umožňují nalezení sekvenčních variant. Přidáním fluorescenčního barviva se však změní podmínky PCR reakce, což je zásadní nevýhoda znemožňující plné převzetí PCR podmínek z již zavedených systémů. HRM analýza začíná zahřátím DNA fragmentů z teploty 55 °C na teplotu 90°C. Někde v tomto rozmezí se nachází teplota tání DNA fragmentu (T_m), což je teplota, při které je denaturována polovina fragmentů. Polovina DNA je tedy ve formě dvoušroubovice a druhá polovina ve formě jednovláknové DNA (obrázek 1.4). Hodnota T_m je velmi závislá na pořadí bází. Čím více G-C párů obsahuje daný fragment, tím větší je teplota tání, což je způsobeno větším počtem vodíkových vazeb mezi G-C páry v porovnání s A-T páry [21]. Na konci fáze tání jsou ve vzorku přítomné jen molekuly jednovláknové DNA (ssDNA). Pro jednotlivé vzorky vypočítává hodnota skóre a rozlišení a určí se podobnost vzorku s příslušným standardem.



Obrázek 1.4 HRM křivka (upraveno z [22])

2 Metody bakteriální typizace

Pojem typizace označuje diferenciaci druhů na nižší taxony na základě genetické variability. Přitom podle charakteru detekovaného znaku můžeme metody typizace rozdělit na fenotypové a genotypové. První metody, které umožňovaly bakterie typizovat, zkoumaly především fenotypové charakteristiky jednotlivých kmenů. V 90. letech byly izoláty bakterií rozlišovány na základě sérotypizace [23] nebo rozdílné citlivosti k antibiotikům [24]. Dnes se díky vyšší rozlišovací schopnosti prosazují především molekulárně-genetické metody rozlišující kmeny dle jejich genetické variability [25]. Těchto metod dnes existuje velké množství a zvolení nejvhodnější metody není úplně jednoduchý úkol. Je potřeba vybírat nejen s ohledem na náročnost a rychlost metody, ale také s ohledem na vybavení laboratoře [26]. Pro přehlednost můžeme genotypizační techniky rozdělit podle jejich hlavního principu na elektroforetické, založené na PCR a založené na sekvenaci DNA.

Mezi nové a rozvíjející se metody založené na sekvenaci DNA můžeme zařadit typizaci dle variability sekvence genu *slpA* (gen kódující protein A nacházející se na vnější straně buněčné membrány) [27] a typizaci dle sekvencí mnoha lokusů (MLST) [28]. Do druhé skupiny patří především pulzní gelová elektroforéza (PFGE) [29] restrikční endonukleázovou analýzu (REA) [30] a polymorfismus délky restrikčních fragmentů (RFLP) [31]. Mezi metody založené na PCR můžeme zařadit multilokusovou analýzu tandemových repetitivních sekvencí (MLVA) [32], polymorfismus délky amplifikovaných fragmentů (AFLP) [33], ribotypizaci [34], toxinotypizaci [35], náhodně amplifikovanou polymorfní DNA (RAPD) [36] a pro tuto práci zásadní metodu mini typizaci dle sekvencí mnoha lokusů (mini-MLST) [29].

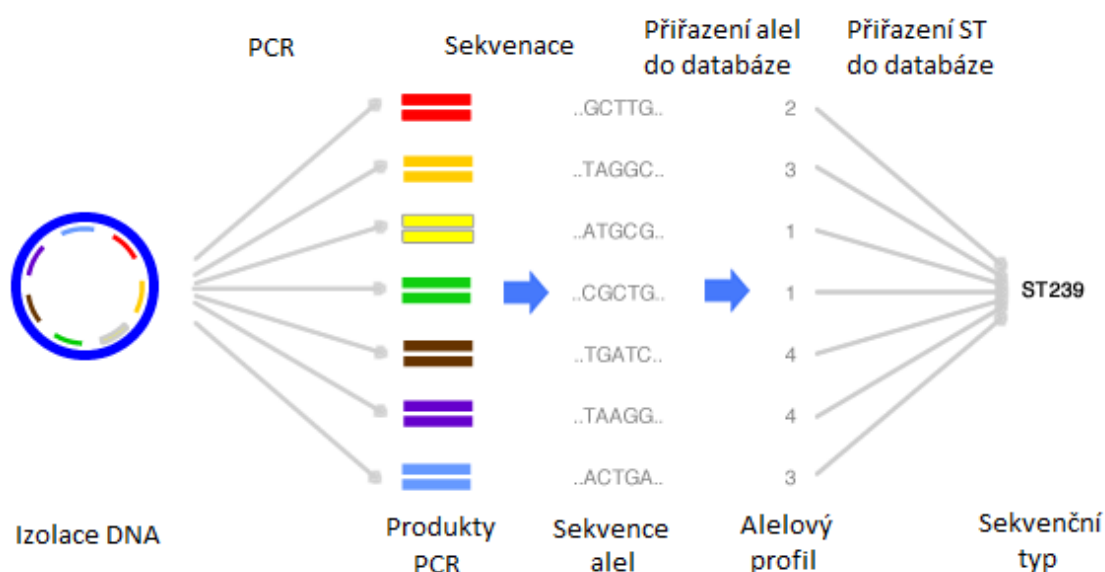
2.1 Typizační metody založené na sekvenaci DNA

Typizace dle variability sekvence genu *slpA*

Tato metoda je známá pod anglickým označením Surface Layer Protein A Gene Sequence Typing (zkráceně *slpAST*). Gen *slpA* (Surface Layer Protein A) kóduje dva druhy proteinů nacházejících se na vnější straně buněčné stěny. Protein o větší molekulové hmotnosti (vzniklý z C-konce prekurzoru) je vysoce konzervovaný, sekvence druhého proteinu (z N-konce prekurzoru) je značně variabilní, což umožňuje bakterie odlišovat [27]. Následnou amplifikací je získán produkt délky 0,5-1 kb, který je dále sekvenován. Od něj je odvozeno pořadí aminokyselin a bakterie je zařazena. Výhodou je schopnost provést zatřídění přímo z klinických vzorků, čímž odpadá potřeba procesu kultivace bakterií. Dále je výhodou vysoká rozlišovací schopnost a snadná reprodukovatelnost dat [37].

Typizace dle sekvencí mnoha lokusů

Tato metoda je anglicky označovaná jako Multi Locus Sequence Typing, zkráceně MLST. Principem je typizace na základě amplifikace a následné sekvenace úseků vhodně zvolených referenčních nebo také provozních (anglicky housekeeping) genů. Obvykle se volí sedm a více lokusů a délka fragmentů je v rozmezí 400-700 bp. Nejprve je potřeba připravit DNA vzorky, následně probíhá PCR amplifikace pomocí primerů. Ty jsou navrženy ze sekvencí homologních genů z gram-pozitivních bakterií s nízkým obsahem cytosinu a guaninu. Produkty PCR jsou následně purifikovány a sekvenovány s PCR dopřednými nebo reverzními primery. Výsledky se analyzují a porovnávají se sekvencemi uloženými v databázích. Každý analyzovaný úsek je přiřazen k jedné alele definované pořadovým číslem v databázi [28]. Jakákoli odlišnost (třeba jen jedné báze) v sekvenci daného provozního genu z něho dělá novou alelu. Výsledkem je tedy číselný kód podle počtu použitých provozních genů, takzvaný alelický profil, ze kterého se následně určuje sekvenční typ. Postup je zobrazen na obrázku 2.1. Následně se ještě porovná podobnost jednotlivých sekvenčních typů a vytvoří se dendrogram. MLST má několik výhod oproti jiným metodám. Data (DNA sekvence) jsou jednoznačná, takže je lze snadno srovnávat mezi různými laboratořemi a mohou být uložena ve sdílené centrální databázi, aby poskytla širší zdroj pro epidemiologické studie [38], např. <http://www.mlst.net/>, <http://www.pasteur.fr/>, <https://pubmlst.org/> (obrázek 2.2), Dále také vykazuje vyšší citlivost, která je způsobena schopností detekovat a indexovat neutrální genetické variace (ty se v rámci nukleotidových sekvencí provozních genů akumulují poměrně pomalu a jsou tedy vhodné pro fylogenetickou analýzu) [39]. I přesto, že je tato metoda velmi efektivní, vyžaduje 14 samostatných sekvenačních reakcí na jeden izolát. Z časových a finančních důvodů ji tedy nelze využít v běžném laboratorním provozu a nachází uplatnění pouze při rozsáhlých výzkumných projektech [40].



Obrázek 2.1 Postup MLST (upraveno z [41])

Databases hosted on PubMLST

These databases host MLST schemes and isolate data, increasingly including whole genome sequences.

Bacteria

- | | | | |
|--|--|--|---|
| <ul style="list-style-type: none"><i>Achromobacter</i><i>Acinetobacter baumannii</i><i>Aeromonas</i> spp.<i>Anaplasma phagocytophilum</i><i>Arcobacter</i> spp.<i>Bacillus cereus</i><i>Bacillus licheniformis</i><i>Bacillus subtilis</i><i>Bordetella</i> spp.<i>Borrelia</i> spp.<i>Bartonella bacilliformis</i><i>Bartonella henselae</i><i>Brachyspira</i> spp.<i>Brucella</i> spp.<i>Burkholderia cepacia</i> complex<i>Burkholderia pseudomallei</i><i>Campylobacter</i> spp.<i>Carnobacterium maltaromaticum</i><i>Chlamydiales</i> spp.<i>Citrobacter freundii</i><i>Clostridium botulinum</i><i>Clostridium difficile</i><i>Clostridium septicum</i><i>Corynebacterium diphtheriae</i><i>Cronobacter</i> spp.<i>Dichelobacter nodosus</i> | <ul style="list-style-type: none"><i>Enterobacter cloacae</i><i>Edwardsiella</i> spp.<i>Enterococcus faecalis</i><i>Enterococcus faecium</i><i>Flavobacterium psychrophilum</i><i>Gallibacterium anatis</i><i>Haemophilus influenzae</i><i>Haemophilus parasuis</i><i>Helicobacter cinaedi</i><i>Helicobacter pylori</i><i>Helicobacter suis</i><i>Klebsiella aerogenes</i><i>Klebsiella oxytoca</i><i>Lactobacillus salivarius</i><i>Leptospira</i> spp.<i>Macrococcus canis</i><i>Macrococcus caseolyticus</i><i>Mannheimia haemolytica</i><i>Melissococcus plutonius</i><i>Mycobacteria</i> spp.<i>Mycobacterium abscessus</i> complex<i>Mycoplasma agalactiae</i><i>Mycoplasma bovis</i><i>Mycoplasma hyopneumoniae</i><i>Mycoplasma hyorhinis</i><i>Mycoplasma iowae</i> | <ul style="list-style-type: none"><i>Mycoplasma pneumoniae</i><i>Mycoplasma synoviae</i><i>Neisseria</i> spp.<i>Oral Streptococcus</i> spp.<i>Orientia tsutsugamushi</i><i>Ornithobacterium rhinotracheale</i><i>Paenibacillus larvae</i><i>Pasteurella multocida</i><i>Pediococcus pentosaceus</i><i>Photobacterium damsela</i><i>Porphyromonas gingivalis</i><i>Propionibacterium acnes</i><i>Pseudomonas aeruginosa</i><i>Pseudomonas fluorescens</i><i>Rhodococcus equi</i><i>Riemerella anatipestifer</i><i>Sinorhizobium</i> spp.<i>Staphylococcus aureus</i><i>Staphylococcus epidermidis</i><i>Staphylococcus haemolyticus</i><i>Staphylococcus hominis</i><i>Staphylococcus pseudintermedius</i><i>Stenotrophomonas maltophilia</i><i>Streptococcus agalactiae</i><i>Streptococcus bovis/equinus</i> complex<i>Streptococcus canis</i> | <ul style="list-style-type: none"><i>Streptococcus dysgalactiae</i><i>Streptococcus gallolyticus</i><i>Streptococcus pneumoniae</i><i>Streptococcus pyogenes</i><i>Streptococcus suis</i><i>Streptococcus thermophilus</i><i>Streptococcus uberis</i><i>Streptococcus zooepidemicus</i><i>Streptomyces</i> spp.<i>Taylorella</i> spp.<i>Tenacibaculum</i> spp.<i>Vibrio</i> spp.<i>Vibrio cholerae</i><i>Vibrio parahaemolyticus</i><i>Vibrio tapetis</i><i>Vibrio vulnificus</i><i>Wolbachia</i> spp.<i>Xylella fastidiosa</i><i>Yersinia pseudotuberculosis</i> (legacy)<i>Yersinia</i> spp.<i>Yersinia ruckeri</i> |
|--|--|--|---|

Obrázek 2.2 Přehled veřejných databází PubMLST (převzato z [42])

2.2 Typizační metody založené na elektroforéze

Pulsní gelová elektroforéza

Pulsní gelová elektroforéza (anglicky Pulsed-Field Gel Electrophoresis, odtud zkratka PFGE) je často považována za zlatý standard mezi typizačními metodami. Patří mezi nejužívanější postupy k určení izolátů bakterií při rozsáhlých epidemiologických studiích [29]. Metoda vychází z klasické agaróзовé elektroforézy, od které se odlišuje periodickými změnami orientace elektrického pole. Tak dojde k reorientaci molekuly, tzn. uvedení molekuly do stavu, kdy je schopna dalšího pohybu. Čím je molekula menší, tím rychleji se reorientuje (má kratší reorientační čas) a výsledný přímočarý pohyb je rychlejší. Výhodou je, že na rozdíl od klasické elektroforézy umožňuje tato metoda efektivní dělení molekul DNA o velikosti 15 kb až 10 Mb. Cílem je naštěpení celého bakteriálního genomu pomocí specifické restrikční endonukleázy na několik kratších fragmentů a provedení analýzy restrikčního profilu [30]. Hlavní výhodou PFGE je její vysoká diskriminační síla [43], poměrně levný postup a výborná schopnost zatřídění. Naopak nevýhodou je technická časová náročnost metody.

Restrikční endonukleázová analýza

Restrikční endonukleázová analýza (anglicky Restriction Endonuclease Analysis, zkráceně REA) je metoda založená na štěpení bakteriálního genomu speciálním enzymem pracujícím s mnohem větší frekvencí, než je tomu u metody PFGE. Výstupem jsou tedy stovky krátkých fragmentů separovaných za pomoci klasické gelové elektroforézy dosahujících maximálně délky 30 kb [30]. Vizuální vyhodnocení stovek fragmentů je v tomto případě příliš obtížné. Profil vzorku je tedy nanesen na elektroforetický gel spolu se vzorovými kmeny a po milimetrových segmentech porovnáván. Výsledný restrikční profil se vyhodnocuje výpočtem tzv. indexu podobnosti (SI), což je procentuální zastoupení identických segmentů z 60 analyzovaných. Pokud je shoda s profilem některého vzorového kmene 100 %, je izolát

zařazen. Pokud je shoda se všemi vzorovými kmeny menší než 90 %, izolát je zařazen do nové, samostatné skupiny [44]. Výhodou této typizační techniky je její vysoká rychlost, rozlišovací schopnost a citlivost. Nevýhodou je těžko proveditelná mezilaboratorní výměna dat, což je způsobeno přímého porovnávání vzorků jednomu gelu.

Polymorfismus délky restrikčních fragmentů

Polymorfismus délky restrikčních fragmentů, též označovaný jako restrikční analýza (anglicky Restriction Fragment Length Polymorphism, zkráceně RFLP) je technika využívající restrikčních endonukleáz ke štěpení DNA. Restrikční endonukleáza rozpozná a naštěpí úsek DNA, který je dlouhý jen několik párů bazí. Pokud jedna alela určitého genu obsahuje rozpoznávací sekvenci, zatímco jiná nikoliv, bude DNA první alely rozštěpena, zatímco DNA druhé alely zůstane vcelku (obrázek 2.3). Dochází ke vzniku různě dlouhých fragmentů a jejich odlišné rychlosti migrace při elektroforéze. Jednotlivé endonukleázy se liší nejen požadavky na optimální podmínky, ale i místy restrikce, např. restriktáza označovaná jako *EcoRI* rozpozná v sekvenci **5'GAATTC3'** a rozštěpí ji v místě **5'G|AATTC3'**. Výhodou metody je výborná reprodukovatelnost a interpretovatelnost [45]. Mezi nevýhody pak patří velké náklady, rizika práce s radioaktivním materiálem, potřeba velkého množství DNA a kvalifikace personálu.

a)

5'ATTAGCCTTCAGTTCGAAAGTAGCT**G|AATTC**GTCAGCTTTCGAAAGTAGCT**G|AATTC**GTCA3'

b)

5'ATTAGCCTTCAGTTCGAAAGTAGCT**GACTTC**GTCAGCTTTCGAAAGTAGCT**G|AATTC**GTCA3'

c)

5'ATTAG**G|AATTC**AGTTCGAAAGTAGCT**G|AATTC**GTCAGCTTTCGAAAGTAGCT**G|AATTC**GTCA3'

Obrázek 2.3 Restrikční analýza a) dvě restrikční místa (3 fragmenty), b) první restrikční místo zrušeno záměnou nukleotidu A na C (2 fragmenty), c) vznik nového restrikčního místa záměnou nukleotidů CC na AA (4 fragmenty)

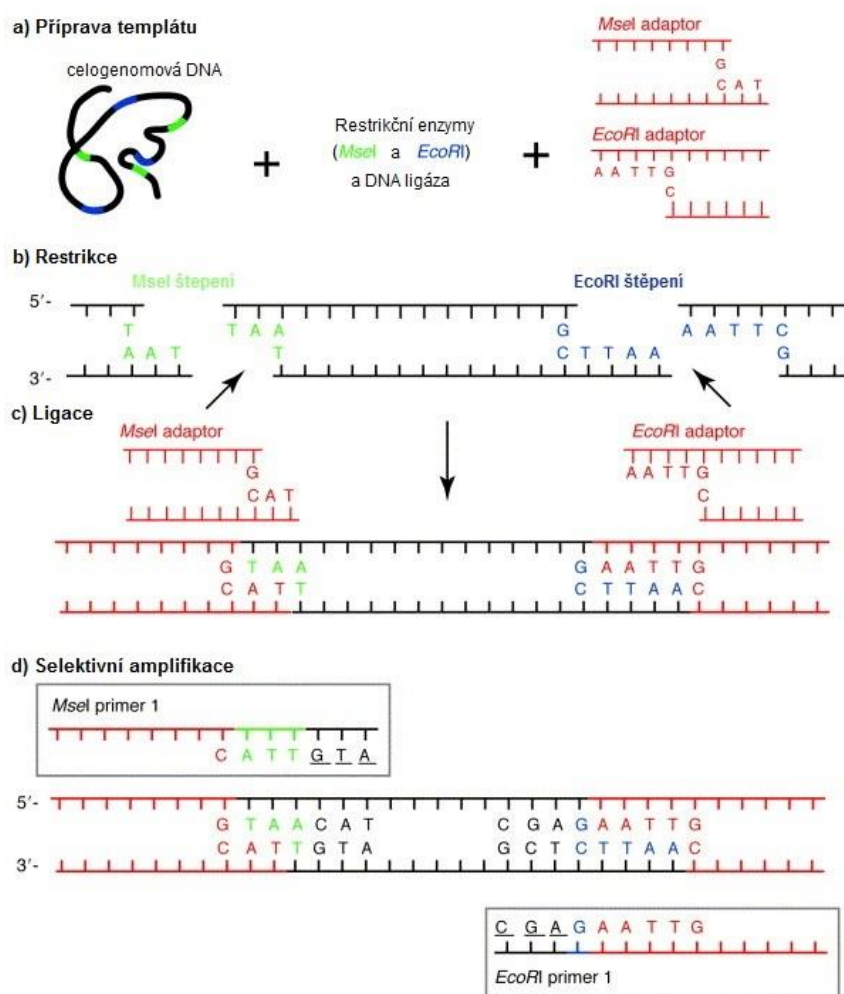
2.3 Typizační metody založené na PCR

Multilokusová analýza tandemových repetitivních sekvencí

Multilokusová analýza tandemových repetitivních sekvencí (anglicky Multi-Locus Variable Number Tandem Repeat Analysis, zkráceně MLVA) dokáže rozlišit kmeny podle rozdílů v počtu tandemových repetitivních sekvencí v několika definovaných lokusech v genomu [32]. Analyzovány mohou být pouze lokusy obsahující krátké vysoce konzervované repetitivní úseky s velkým počtem opakování, což většina lokusů, přestože má repetitivní sekvence, nesplní. Dalšími požadavky je dostatečná variabilita a krátký specifický produkt amplifikace těchto lokusů. Díky těmto podmínkám bylo např. k typizaci *C. difficile* nalezeno pouze sedm vhodných lokusů [46]. Počty tandemových repetitivních sekvencí jsou v 28 těchto vybraných lokusech stabilní a do jedné podskupiny MLVA můžeme zařadit kmeny lišící se maximálně jednou repetitivní sekvencí. Výhodou je vysoká rozlišovací schopnost. MLVA dokáže rozeznat desítky různých podtypů v rámci jednoho ribotypu [47].

Polymorfizmus délky amplifikovaných fragmentů

Polymorfizmus délky amplifikovaných fragmentů (anglicky Amplified Fragment Length Polymorphism, zkráceně AFLP) je vhodná metoda pro populační studie. Kombinuje výhody principu PCR a štěpení restrikční endonukleázou a dokáže zjistit vztahy na úrovni blízkce příbuzných taxonů. Samotná metoda má několik fází (obrázek 2.4) [33]. Během restrikce se specificky štěpí celková DNA pomocí dvou restrikčních endonukleáz *MseI* a *EcoRI*. *MseI* rozpoznává sekvenci o délce 4bp „TTAA“ a *EcoRI* 6bp sekvenci „GAATTC“. Na konci fáze restrikce dostaneme tedy velké množství fragmentů, které mají na jednom konci *MseI* a na druhém *EcoRI*. Během fáze ligace jsou ke všem fragmentům připojeny adaptory. Dále probíhá prosektivní amplifikace, kdy známé počáteční sekvence všech fragmentů můžeme namnožit pomocí specifických primerů a polymerázové řetězové reakce. Posledním krokem je selektivní amplifikace a redukce počtu fragmentů pro spolehlivé vyhodnocení. Technika je snadná, rychlá a všestranná. Nemá tedy potřebu znalostí studovaného organismu. Omezení spočívá ve vysokých finančních nákladech, požadavku na kvalitní DNA a nižší rozlišovací schopnosti podobně jako u ribotypizace [48].



Obrázek 2.4 Fáze metody polymorfismu délky amplifikovaných fragmentů (upraveno z [33])

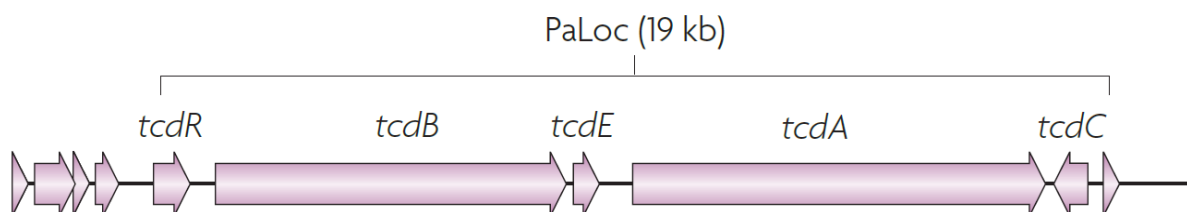
PCR ribotypizace

PCR ribotypizace je metoda při níž jsou amplifikovány variabilní sekvence DNA, které se nacházejí mezi vysoce konzervovanými geny kódujícími 16 S a 23 S rRNA. Na základě rozdílů v délce, sekvenci a počtu alel těchto variabilních oblastí jsou potom jednotlivé kmeny rozděleny do ribotypů. Každý ribotyp pak tvoří specifický elektroforetický profil, podle kterého může být identifikován [34].

Nevýhodou metody je nedostatečné množství standardizovaných sbírek typových kmenů, což brání porovnávání výsledků a tím vede ke vzniku nejednotné nomenklatury. Další nevýhodou je nízká rozlišovací schopnost, a to obzvláště u velkých fragmentů DNA. Tuto nevýhodu však lze eliminovat tak, že klasickou PCR ribotypizaci modifikujeme. K separaci zmíněných variabilních oblastí lze použít kapilární gelovou elektroforézu, což umožní rozlišit mezi fragmenty DNA i rozdíly jednoho nukleotidu. Tato metoda je oproti klasické elektroforéze na agarovém gelu nejen přesnější, ale výstupem jsou numerická data, která je možno nahrávat do databází a dále porovnávat [49]. Již existuje internetová databáze ribotypů generovaných pomocí kapilární gelové elektroforézy (<http://webribo.ages.at>), která pomáhá uživatelům s identifikací známých ribotypů, popřípadě jim umožňuje vkládat ribotypy nové.

Toxinotypizace

Toxinotypizace je další z metod řadících se mezi PCR. Jedná se o analýzu polymorfismu určité oblasti bakteriálního chromozomu označované jako lokus patogenicity (zkráceně PaLoc). DNA sekvence oblasti PaLoc je značně variabilní. Například pro bakterii *Clostridium difficile* exotoxiny *TcdA* a *TcdB* jsou hlavní virulentní faktory produkované patogenními kmeny. Geny kódující tyto exotoxiny se nachází právě na PaLoc [35] (obrázek 2.5). Dá se říci, že tato metoda je spojením amplifikace variabilních oblastí lokusu PaLoc pomocí PCR a analýzy amplifikovaných fragmentů DNA štěpených specifickými endonukleázami (RFLP) [50].



Obrázek 2.5 PaLoc kódující dva geny pro bakteriální toxiny, *tcdA* a *tcdB*, a tři regulační geny, *tcdR*, *tcdE* a *tcdC* [35]

Náhodně amplifikovaná polymorfní DNA

Náhodně amplifikovaná polymorfní DNA (anglicky Random Amplification of Polymorphic DNA, odtud zkratka RAPD) je metoda využívaná především k určování variability na populační úrovni, například k definici klonů. Dnes je k dispozici pro tuto metodu 40 dekanukleotidových primerů. Princip metody je založený na vytvoření příslušného PCR produktu. V případě existence dvou identických (nebo velmi podobných) míst pro nasednutí primeru na reverzně komplementární vlákna DNA, kdy amplifikovatelná vzdálenost mezi dvěma místy je zpravidla do 3000 bp. Následně se provádí elektroforetická vizualizace RAPD variability [36].

Hlavní předností metody je její rychlost a jednoduchost a schopnost odrážet variabilitu skrz celý genom [51]. Dále pak fakt, že není potřeba žádná předchozí molekulární znalost o zkoumaném taxonu, stejně ani velkého množství DNA k analýze – postačuje cca 25 ng. Úskalí metody spočívá v citlivosti RAPD k mnoha parametrům. Naráží se na problém s reprodukovatelností a porovnatelností výsledků není možné srovnání mezi různými studiemi.

Mini typizace dle sekvencí mnoha lokusů

Mini typizace (anglicky Minim Typing nebo také mini-MLST) je efektivní metoda odvozená z principu MLST. Místo sekvenování však využívá analýza jednonukleotidových polymorfismů v rámci úseků MLST lokusů s využitím analýzy HRM [29]. Metoda využívá přístup *in silico*, což v překladu znamená spočteno počítačem. Klíčové prvky této metody jsou seznam jednonukleotidových polymorfismů (zkráceně SNP) nalezených v databázi MLST, PCR amplifikovatelné fragmenty a klíč umožňující překlad dat HRM analýzy do MLST databáze. Nejprve proběhne hledání oblastí s maximální diverzitou v genech pro MLST a navržení primerů amplifikujících části těchto genů. Tyto úseky, tzv. minim fragmenty, jsou velké přibližně 50-150 bp [40]. Ihned po PCR je spuštěna HRM analýza schopná zachytit změnu jediné báze mezi jinak identickými nukleotidovými sekvencemi. Analýzu je možno provést jak na speciálních HRM přístrojích, tak i na upravených real-time PCR přístrojích, které HRM analýzu umožňují. Podrobnější popis HRM analýzy je uveden v podkapitole 1.3.

Podstatou je sledování změn fluorescence při denaturaci vzniklých ampliconů za zvyšující se teploty. Se vzrůstající teplotou se totiž dvoušroubovice DNA rozpadá a fluorescence se vytrácí. K obarvení se využívají barviva vykazující fluorescenci po zakomponování do dvoušroubovice DNA [21]. Při HRM analýze se sleduje výsledná hodnota (anglicky Melting Temperature) nebo tvar získané křivky tání (obrázek 1.4). Pro jednotlivé alely jsou získány rozdílné křivky tání charakterizované především obsahem G-C párů. Kombinací alel zadaných do převodního klíče lze zjistit ke kterému melt typu kmen náleží a interpretovat výsledky podobně jako MLST databáze. Prozatím byla metoda aplikována pouze pro nejznámější bakterie jako např. *Streptococcus pyogenes*, *Staphylococcus aureus*, *Enterococcus faecium* nebo *Klebsiella pneumoniae*. Přesto, že rozlišovací schopnost mini-MLST je o něco menší než MLST[40], metoda má veliký potenciál. Obrovskou předností této metody jsou pouze pětinové až desetinové náklady oproti MLST díky vynechání kroku sekvenace genů. Výhodou HRM oproti sekvenaci je totiž jednoduchost (nevyžaduje přidávání dalších reagentů), rychlost, kapacita a nižší cena. Doposud však neexistují žádné databáze podobné již zmíněným MLST databázím (obrázek 2.2)

3 Návrh převodního klíče

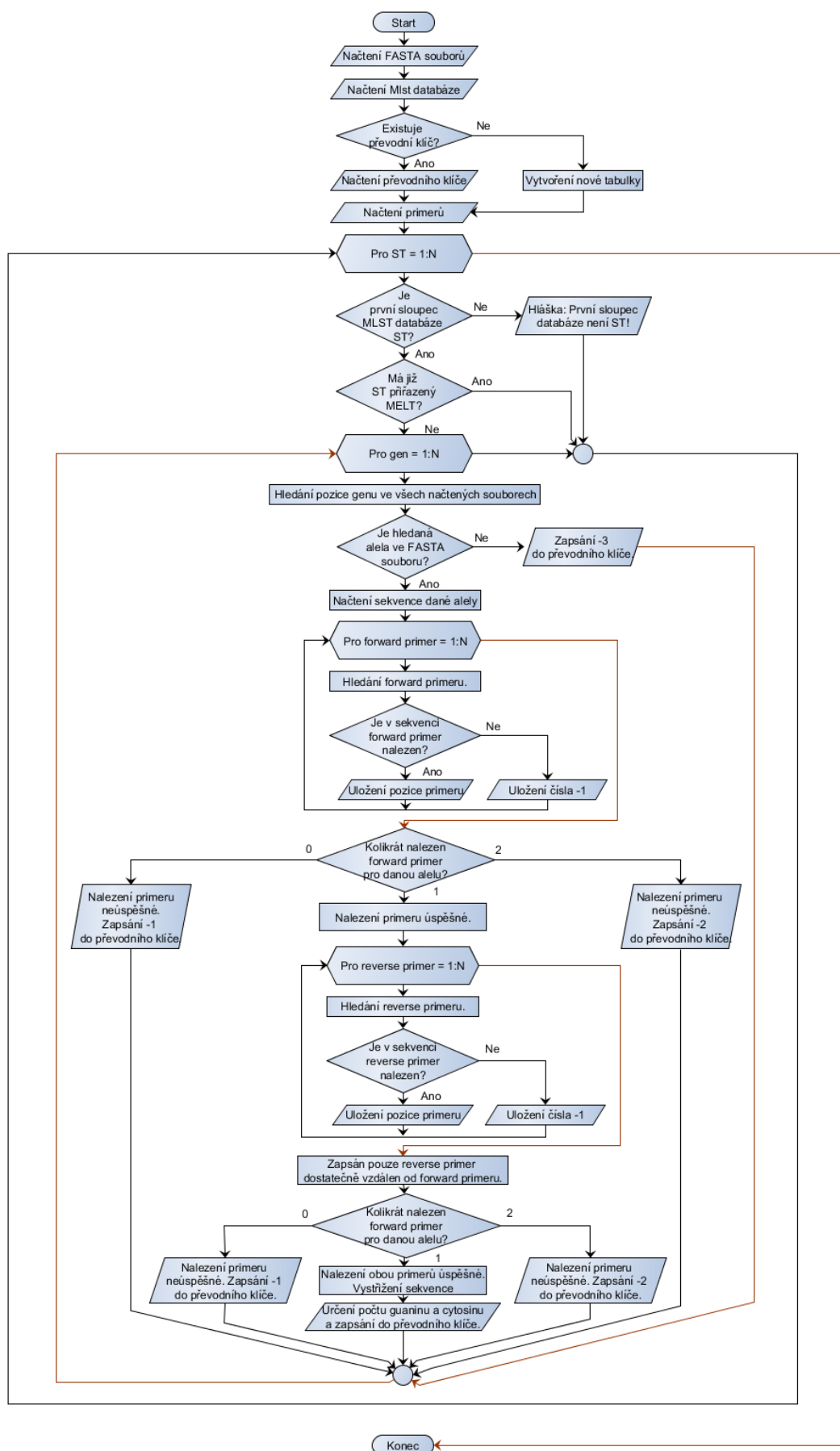
V předchozí kapitole byly podrobně popsány metody, jak lze typizovat patogenní bakterie. Z finančních důvodů se však většinou typizace bakterií využívá spíše pro výzkum než pro rutinní práci v laboratorním provozu. Samotným mikrobiologům sice stačí, že na konkrétní bakterii je třeba dát určité antibiotikum a původ bakterie již znát nepotřebují. Pokud však bude možné typizovat bakterie snadno a levně, bude možné sledovat jejich šíření a předcházet mu. Za tímto účelem se tedy hledalo levné a spolehlivé řešení. Základní převodní klíč byl navržen ve spolupráci s Centrem molekulární biologie a genové terapie Interní hematologické a onkologické kliniky Fakultní nemocnice Brno. Principem metody je překlad sekvenčního typu na příslušející melt typ umožňujícího srovnání sekvenčních a PCR dat. Celý vývojový diagram je zobrazen na obrázcích 3.1 a 3.2.

V prvních krocích je potřeba načíst do proměnných data potřebná k práci s převodním klíčem. Data jsou postupně načítána podle cesty zvolené uživatelem. Po načtení FASTA souborů s alelami genů a MLST databáze je potřeba rozhodnout, zda bude vytvořena nová tabulka převodního klíče nebo aktualizována již existující. Podle rozhodnutí uživatele je pak opět klíč načten podle zvolené cesty, nebo je vytvořena nová prázdná tabulka (například naplněná nulami). Počet řádků tabulky bude udávat počet ST podle MLST databáze plus jeden řádek jako hlavička. Počet sloupců je dán počtem načtených FASTA souborů a dalších dvou sloupců pro ST a MELT typ. Dalším krokem je načtení souboru se všemi možnými variantami forward a reverse primerů pro všechny geny, se kterými pracujeme. Protože však u některých genů může být vzdálenost mezi primery moc malá, je potřeba zavést další proměnnou, tzv. minimální vzdálenost. Tento krok ještě není pevně stanoven. Možností je i omezení naopak, tj. nastavení maximální vzdálenosti. Když jsou načtena všechna data, spustí se výpočet vnitřku výstupní tabulky. Postupně jsou procházeny řádky, což jsou jednotlivé ST. Nejprve ověříme, zda je první sloupec ST. Pokud není, bude uživatel upozorněn hláškou.

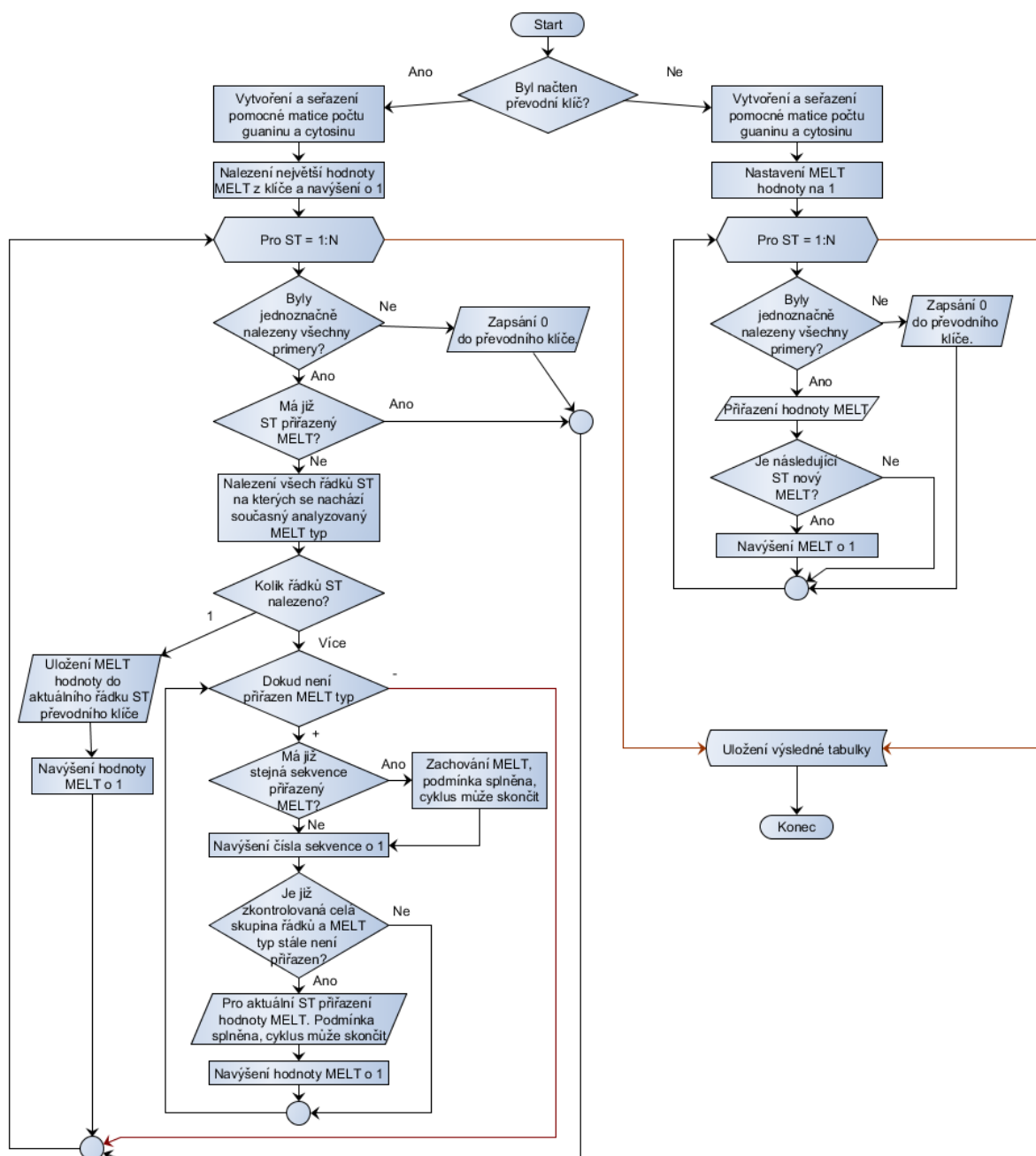
Další ověření, zda už není melt typ přiřazen, je potřeba zajistit pomocí vyhledávání shody libovolného ST v klíči s aktuálním ST v MLST databázi. Pokud melt typ již je přiřazen, celý výpočet je přeskočen a načítá se další ST. Pokud přiřazen není, výpočet se spustí. Pro každý gen (sloupec) MLST databáze je potřeba najít sloupec, kde se nachází daný gen v převodním klíči, kde v matici načtených FASTA souborů a kde v matici s načtenými primery. Dále je potřeba zajistit upozornění uživatele pomocí hlášky, pokud hledaná alela genu v matici načtených FASTA souborů není. Pokud je alela nalezena, načte se její sekvence a začne se hledat forward primer. Postupně jsou brány všechny varianty primeru a vyhledávány v sekvenci. Pokud daná varianta primeru nalezena není, uloží se hodnota -1, v opačném případě je uložena pozice primeru. Takto se vyhledávají všechny varianty primeru, proto je možné, že bude nalezeno více variant nebo naopak žádná. V takovém případě bude uživatel opět

upozorněn. Pokud bude nalezen právě jeden forward primer, spustí se hledání reverse primeru na stejném principu pouze doplněném o podmínku dostatečné vzdálenosti od forward primeru. V případě nalezení právě jednoho forward primeru a právě jednoho reverse primeru je vyříznuta oblast mezi nimi, spočtena suma guaninu a cytosinu a výsledek zapsán do převodního klíče.

V další části bude potřeba přiřadit příslušný melt typ. Oříznutím prvního a posledního sloupce původní matice je vytvořena pomocná matice obsahující pouze počty guaninu a cytosinu, která je následně seřazena podle velikosti. Další krok se liší podle toho, zda byl načten převodní klíč nebo ne. Při aktualizaci převodního klíče je nalezena největší hodnota melt z klíče, která je následně navýšena o jedničku a uložena jako výchozí hodnota při dalším přiřazování melt typu. Pokud se vytváří nový klíč, bude výchozí hodnota melt nastavena na jedničku. Poté je opět procházen každý ST jako řádek matice a probíhá kontrola, zda byly všechny primery nalezeny jednoznačně. Pokud ne, opět je uživatel upozorněn. Pokud se vytváří nový klíč, bude se hodnota melt postupně přiřazovat. Při aktualizaci klíče je nutné dohledat všechny řádky ST na kterých se nachází současný analyzovaný melt typ. V případě nalezení více než jednoho řádku je potřeba pomocí cyklu while všechny řádky v pomocné matici projít a pokud je melt typ přiřazen, tak zapsat do převodního klíče a ukončit cyklus. Jestliže celá skupina nemá přiřazený melt, je do klíče přiřazena nová hodnota, která byla dříve nastavena jako výchozí. Poté je možno cyklus while také ukončit. Výchozí hodnota melt je navýšena o jedničku. Celý postup se pak opakuje pro všechny ostatní ST převodního klíče. Když je výpočet u konce, zvolí uživatel název, pod kterým bude převodní klíč uložen ve formátu .xlsx.



Obrázek 3.1 Vývojový diagram načtení souborů a výpočtu vnitřní části převodního klíče

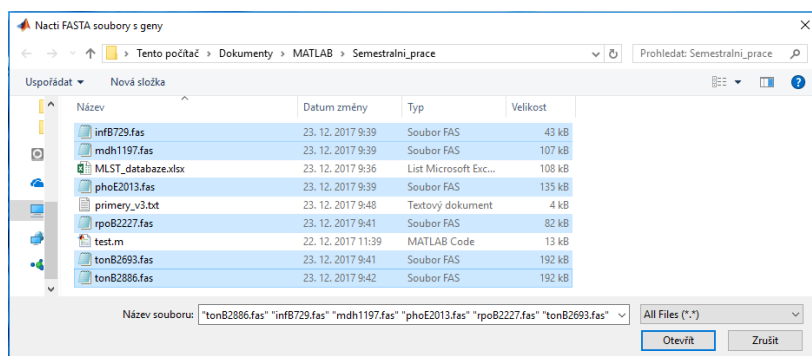


Obrázek 3.2 Vývojový diagram přiřazení MELT typu a uložení tabulky

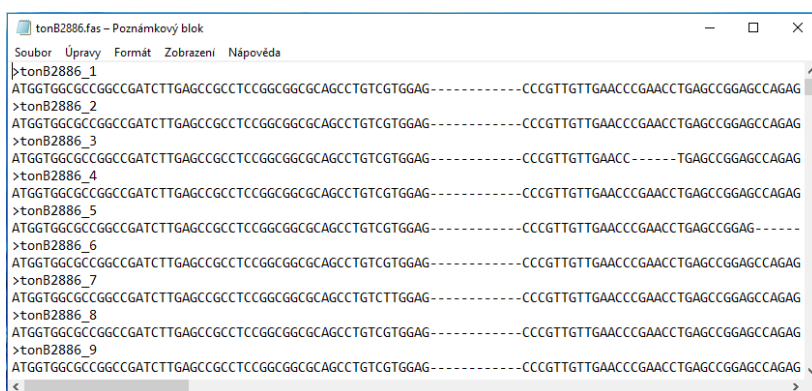
4 Implementace v prostředí MATLAB

Původní převodní klíč byl implementován do prostředí MATLAB. Nejdříve vznikl program bez grafického rozhraní, kterému je věnována tato kapitola. Program byl již otestován a poté i využíván v laboratorním provozu. V následující kapitole je pak popsáno rozšíření programu o uživatelské rozhraní a jeho zakomponování do nástroje pro správu databáze.

Po spuštění původního programu bez uživatelského rozhraní je uživatel dialogovým oknem vyzván k načtení FASTA souborů obsahujících alely genů (obrázek 4.1). Požadovány jsou předem upravené soubory. Sekvence by měly být zarovnané a hlavička každé sekvence musí obsahovat jméno genu včetně velkých/malých písmen, podtržítka a číslo alely (obrázek 4.2). Také počet vybraných genů musí odpovídat počtu genů v databázi a převodním klíči (pokud již existuje). Pro označení více než jednoho souboru je potřeba využít přidržení klávesy Ctrl. Výběr souborů se potvrdí kliknutím na tlačítko Otevřít.



Obrázek 4.1 Dialogové okno Načti FASTA soubory s geny



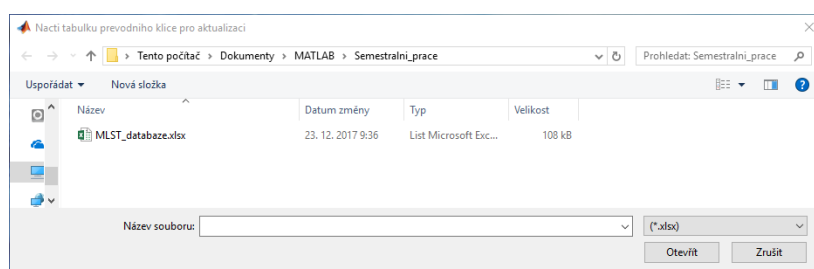
Obrázek 4.2 Ukázka vhodně popsaného a zarovnaného FASTA souboru

Následně je uživatel vyzván dialogovým oknem pro volbu databáze MLST. Databáze musí být exportovaná do excel souboru s jediným listem (případně musí být na prvním listu a další listy musejí být prázdné). Sloupce s geny, které nemají být použity musí být odstraněny. Všechny zbytečné popisky taktéž. První řádek obsahuje jména ST a genů, zbytek pak

kombinace alel (obrázek 4.3). První sloupec musí být ST, pak už na pořadí genů nezáleží. Pokud je nástroj využit k aktualizaci existujícího převodního klíče, ostatní sloupce budou seřazeny tak, jak tomu je v existujícím klíči. Pokud je vytvářen zcela nový klíč, dialogové okno je potřeba zavřít tlačítkem Storno. Bude sestavena zcela nová tabulka pro převod MLST na MELT, kde bude pořadí genů (pořadí sloupců) stejné jako v zadané MLST databázi. Pokud má proběhnout aktualizace již existujícího klíče, je potřeba vybrat soubor se stávajícím klíčem a potvrdit tlačítkem Otevřít (obrázek 4.4). Musí se jednat o excel soubor, opět s jedním listem (případně více listy, ale klíč je na prvním a ostatní jsou prázdné). První řádek obsahuje popis sloupců. První sloupec je ST, následují sloupce genů, poslední sloupec je Melt. Opět je důležité, aby se geny (včetně velkých/malých písmen) shodovaly s geny v databázi a hlavičkách FASTA souborů.

	A	B	C	D	E	F	G
1	ST	infB725	mdh11	phoE20	rpoB22	tonB26	tonB28
2	1	4	1	7	4	10	10
3	2	4	1	9	4	17	17
4	3	5	1	9	6	11	11
5	4	1	1	3	3	1	1
6	5	2	1	3	3	3	3
7	6	10	1	14	3	20	20
8	7	1	1	3	1	34	34
9	8	1	1	1	5	6	6
10	9	1	3	1	2	8	8
11	10	1	7	9	1	12	12
12	11	3	1	1	1	4	4
13	12	3	1	12	1	4	4
14	13	3	1	10	1	19	19
15	14	6	1	1	1	1	1
16	15	1	1	1	1	1	1
17	16	1	2	4	4	4	4

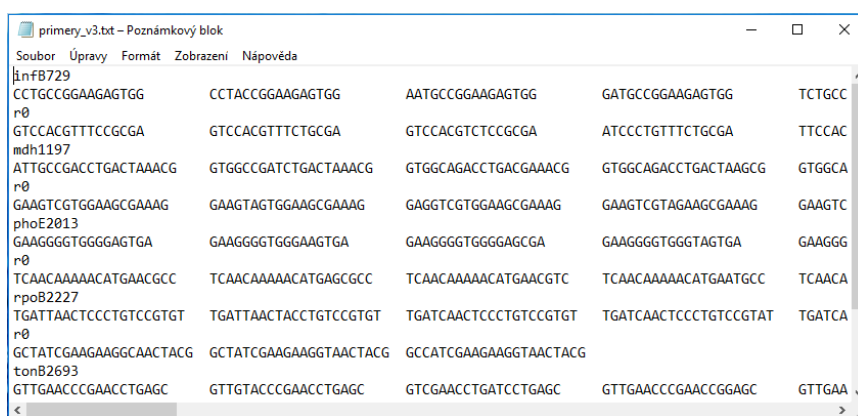
Obrázek 4.3 Ukázka části MLST databáze exportované do excel souboru



Obrázek 4.4 Dialogové okno Načti tabulku převodního klíče pro aktualizaci

Dalším krokem je volba souboru, který obsahuje primery a omezení vzdálenosti mezi primery. Jedná se o textový soubor s přesně danou strukturou, kde se střídají čtveřice řádků, kdy z každé čtveřice vždy první řádek obsahuje jméno genu (zase do písmena stejné jak v databázi, klíči, a FASTA), druhý forward primery oddělené tabulátorem, třetí řádek obsahuje znak „r“ a za ním bez mezery číslo udávající minimální povolenou vzdálenost mezi primery a čtvrtý řádek obsahuje reverse primery opět oddělené tabulátorem. Na pořadí genů nezáleží (obrázek 4.5).

Program následně spustí výpočet. Na posledním dialogovém okně je pak výzva pro definici excel souboru, kam bude uložena vytvořená tabulka. Je potřeba zvolit nový soubor, ne přepisovat starý. Po kliknutí na tlačítko uložit dojde k uložení výsledné excel tabulky a program skončí. Výsledkem je tabulka, kde je každému ST přiřazen MELT (obrázek 4.6). Vnitřek tabulky je absolutní počet cytosinu a guaninu v dané alele. Pokud se však v některé buňce souboru nachází hodnota -1, pro daný ST a určitou alelu nebyl nalezen forward nebo reverse primer. Buňka s hodnotou -2 značí, že bylo nalezeno více primerů a program nedokáže určit výsledný forward nebo reverse primer. Buňka s hodnotou -3 se objeví, pokud ve FASTA souboru nebyla nalezena požadovaná alela. Pokud pak buňka Melt obsahuje hodnotu 0, tak pro daný ST obsahuje nějaký gen hodnotu -1 nebo -2 a nelze tak přiřadit Melt typ.



Obrázek 4.5 Ukázka přesně dané struktury souboru s primery

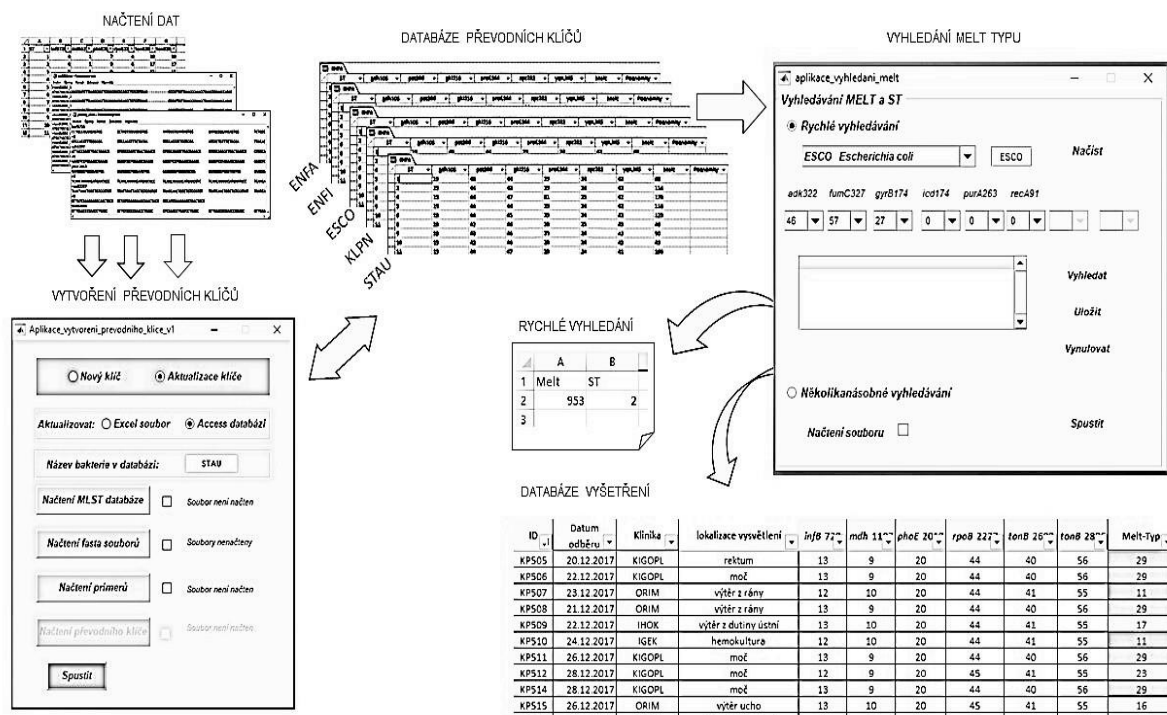
	A	B	C	D	E	F	G	H	I
1	ST	infB729	mdh1197	phoE2013	rpoB2227	tonB2693	tonB2886	Melt	
2	1	12	10	20	45	40	55	65	
3	2	12	10	20	45	41	56	69	
4	3	13	10	20	45	41	55	208	
5	4	13	10	20	43	40	56	190	
6	5	13	10	20	43	-1	56	0	
7	6	13	10	19	43	40	56	148	
8	7	13	10	20	44	41	56	201	
9	8	13	10	20	44	40	55	196	
10	9	13	9	20	45	40	56	118	
11	10	13	9	20	44	41	55	111	
12	11	12	10	20	44	41	55	60	
13	12	12	10	21	44	41	55	73	
14	13	12	10	19	44	40	56	39	
15	14	13	10	20	44	40	56	197	
16	15	13	10	20	44	40	56	197	
17	16	13	9	20	45	41	55	120	

Obrázek 4.6 Ukázka výsledného převodního klíče

5 Nástroj pro správu MS Access databáze

Samotný nástroj je složen ze dvou programů, jejichž popisem se zabývá tato kapitola. Výpočet převodního klíče z předchozích kapitol je součástí prvního z programů. Algoritmus na základě načtených dat (FASTA souborů, souboru s primery a MLST databáze) přiřadí sekvenčním typům příslušný melt typ a umožní uložení klíče buď v podobě tabulky v Excel souboru nebo přímo do Access databáze, kde každá bakterie má svoji tabulku a každý sloupec v tabulce odpovídá jedné konkrétní alele. Data jsou řazena podle prvního sloupce obsahujícího sekvenční typy. Předposlední sloupec obsahuje vždy melt typ příslušný pro daný sekvenční typ. Při načtení nových dat lze klíč i znovu načíst a aktualizovat.

Druhý program je určen pro vyhledávání v databázi. Uživatel zvolí, o kterou bakterii se zajímá, zadá počty G-C párů u zvolených alel a program pak zpětně vyhledává konkrétní melt typ a příslušné sekvenční typy. Tyto výsledky může uživatel pouze zobrazit v programu nebo rovnou uložit do databáze vyšetření pacientů, kde budou sloužit k přesnější identifikaci patogenní bakterie a rychlému a levnému stanovení diagnózy pacienta. Celý tento postup, podobu programů i databází, vstupy a výstupy programů znázorňuje schéma na obrázku 5.1.



Obrázek 5.1 Schéma nástroje

5.1 Propojení prostředí Matlab a Microsoft Access

Na začátku bylo potřeba zvolit, v jakém formátu bude databáze vedena. Po konzultaci ve FN Brno byl zvolen nástroj Microsoft Access. Je součástí balíčku Microsoft Office, takže je velmi snadno dostupný, a zároveň nabízí velké možnosti v oblasti dalšího zpracování dat, jako je vytváření relací, dotazů a dalších aplikací. Hlavním důvodem však byla dobrá znalost nástroje mezi biology samotnými.

Spojení prostředí Matlab s relačními i nerelačními databázemi lze navázat dvěma způsoby. ODBC (Open Database Connectivity) je standardní rozhraní Microsoft Windows®, které umožňuje komunikaci mezi systémy správy databáze a aplikacemi typickými pro C nebo C++. JDBC (Java Database Connectivity) naopak umožňuje komunikaci s aplikacemi napsanými v Oracle Java. Database Toolbox™ má jak knihovnu C++, která umožní připojení k ovladači ODBC, tak i knihovnu Java, která se připojí k ovladači JDBC. Připojení k MS Access je možné pouze přes rozhraní ODBC. Jiné není ze strany MS Access podporováno. Před propojením je potřeba vytvořit databázi s prázdnými tabulkami, které budou obsahovat pouze hlavičky značící alely. Další podmínkou funkčního provozu je také vytvoření cesty k nové databázi přes ovladač ODBC Data Source Administrátor.

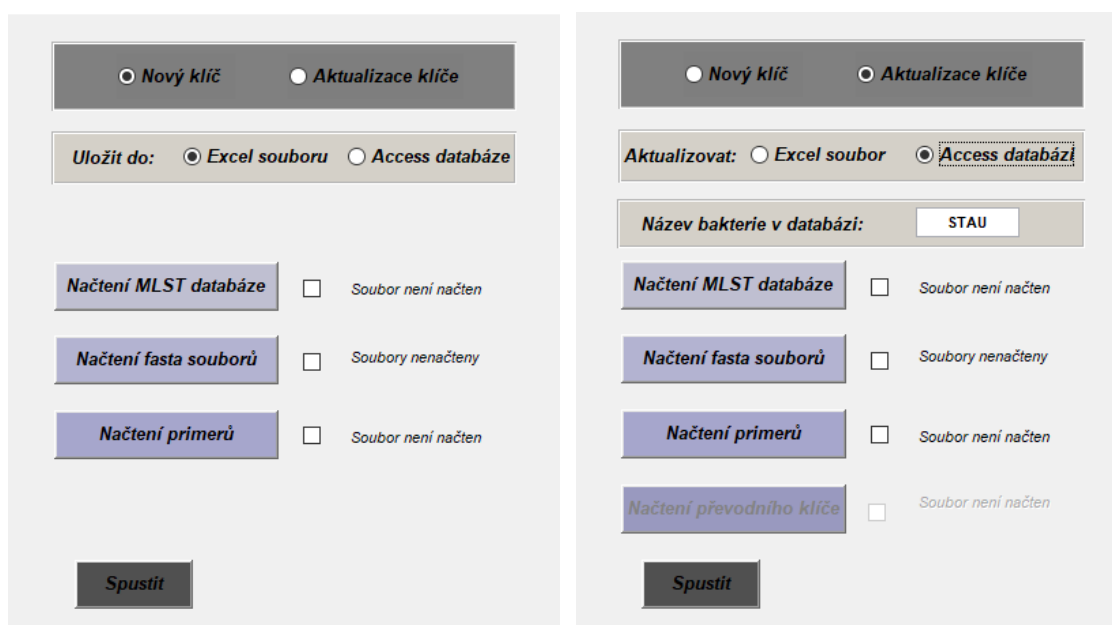
V prostředí Matlab slouží pro výměnu dat s relačními i nerelačními databázemi databázová sada nástrojů (anglicky Database Toolbox), kterou je potřeba do prostředí Matlab nainstalovat. Pro načítání dat je použito několik funkcí. Funkce *database* umožní přístup k souboru a funkce *setdbprefs* zajistí načítání dat v datovém typu cell (česky buňky). Funkcí *select* lze zúžit výběr načítaných dat a funkce *exec* a *fetch* z údajů funkce *database* a *select* vytvoří kompletní cestu a načtou data do struktury. Využít lze také funkci *columnnames*, která umožní načíst hlavičku tabulky. Pro export dat lze využít funkce *insert* pro vložení na nový řádek nebo funkce *update* pro nahrazení dat v tabulce.

Připojení k MS Access je možné pouze přes rozhraní ODBC. Jiné není podporováno. ODBC sice umožňuje rychlejší přenos dat, nemá však takové možnosti jako JDBC. Největší nevýhodou je požadavek na uvedení pevné cesty v přístupu k datům. Znamená to, že k MS Access databázi můžeme přistoupit pouze za podmínky, že uvedeme přesnou cestu k souboru, přesný název souboru/databáze, přesný název tabulky, což je konkrétní bakterie, a dokonce potřebujeme dopředu znát i názvy sloupců označujících alely. Není tedy možné vzdáleně vytvářet nové tabulky ani sloupce. Před použitím programu je tedy nutné příslušnou prázdnou tabulku s hlavičkou vytvořit.

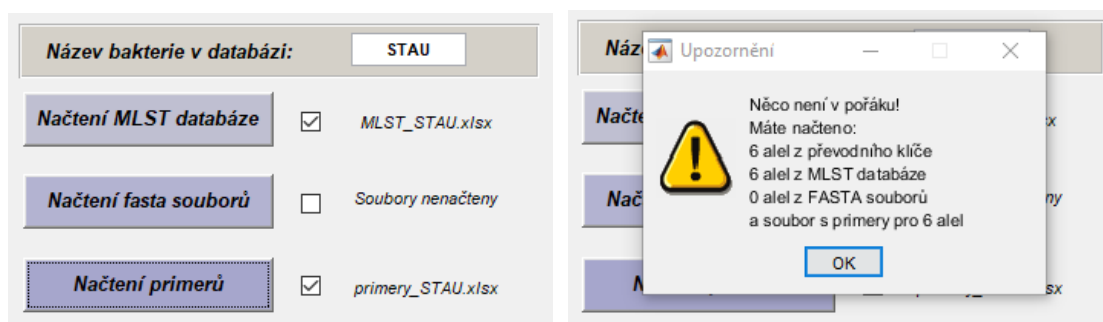
5.2 Program pro správu MS Access databáze

Podoba grafického uživatelského rozhraní

Po spuštění programu se otevře uživatelské rozhraní zobrazené na obrázku 5.2. Rozhraní umožňuje uživateli volbu mezi vytvořením nového klíče nebo aktualizací stávajícího. Druhou volbou je vstupní a výstupní formát. Podle těchto voleb se upraví počet komponent rozhraní, aby měl program všechna data potřebná pro výpočet klíče (obrázek 5.2). Následně uživatel pomocí tlačítek data načte. Každý soubor musí splňovat všechny náležitosti uvedené v kapitole 4. Pro snížení chybovosti je nástroj vybaven po načtení souboru zpětnou vazbou (obrázek 5.3), která uživateli zobrazí název již načteného souboru. Tím se sníží riziko záměny souborů pro jiné bakterie a podobně. Pokud i přesto uživatel načte některé soubory chybně, bude hláškami upozorněn, že něco není v pořádku (obrázek 5.3). Výstupem programu je tabulka převodního klíče ve zvoleném formátu. Každý sloupec v tabulce odpovídá jedné konkrétní alele. Data jsou řazena podle prvního sloupce obsahujícího sekvenční typy. Předposlední sloupec obsahuje vždy melt typ příslušný pro daný sekvenční typ (obrázek 4.6).



Obrázek 5.2 Uživatelské rozhraní a) po spuštění programu b) při změně nastavení



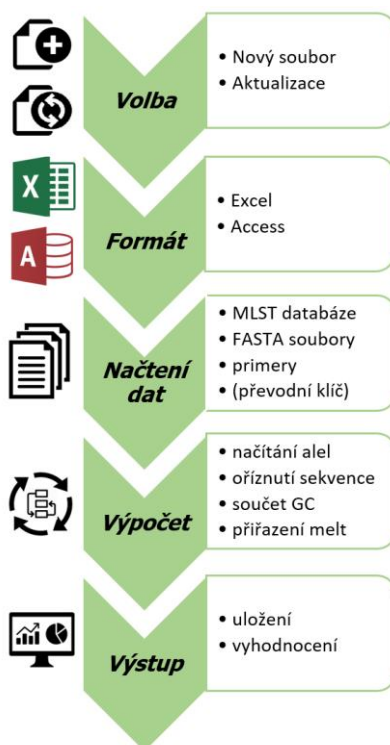
Obrázek 5.3 Zpětná vazba programu a) zobrazení načtených souborů b) upozornění na chybu

Průběh programu a změny v původním převodním klíči

Program obsahuje výpočet převodního klíče kapitoly 3 s menšími úpravami. Testování v laboratorním provozu ukázalo na možná vylepšení, která byla následně aplikována. První změna byla provedena kvůli přehlednosti a snadnější manipulaci s primery. Formát .txt byl nahrazen formátem .xlsx, kde každý list značí jednu alelu. Nová podoba souboru s primery je zobrazena na obrázku 5.4. Jak již bylo řečeno, algoritmus na základě načtených dat přiřadí sekvenčním typům příslušný melt typ a umožní uložení klíče buď v podobě tabulky v Excel souboru nebo MS Access databáze, což je další ze zásadních změn v původním převodním klíči. Schéma programu je zobrazeno na obrázku 5.5 a podoba MS Access databáze na obrázku 5.6. Další změnou je doplnění tabulky o sloupec určený na poznámky. Melt typ příslušný pro daný sekvenční typ je tedy uveden na předposledním sloupci. Změněny byly i některé datové typy. Nyní program pro většinu dat využívá číselný datový typ double. Výjimku tvoří pouze hlavička a sloupec pro poznámky, které mají podobu strings (česky textového řetězce). Rozdílné datové typy však komplikují ukládání výsledků a musí být nakonec před uložením ještě převedeny na společný datový typ cell array (česky buňková pole).

	A	B	C	D
1	GTTGAACCCGAACCTGAGC	TAAGCCGAAGCCCAAACC	50	
2	GTTGTACCCGAACCTGAGC	GAAGCCGAAGCCTAAACC		
3	GTCGAACCTGATCCTGAGC	GAAGCCGAAGCCGAAGCC		
4	GTTGAACCCGAACCGGAGC	GAAGCCGAAGCCAAACCC		
5	GTTGAACCCGAACCAAGAGC	TAAGCCGAAGCCAAACCT		
6	GTTGAACCTGAACCGGAGC	GAAGCCGAAGCCAAACCA		

Obrázek 5.4 Nový zápis primerů v Excel formátu



Obrázek 5.5 Schéma programu pro správu databáze

ST	gdh108	pst566	gki216	aroE344	xpt332	yqiL365	Melt	Poznámky
1	19	43	44	29	24	42	66	
2	19	44	43	29	24	42	114	
4	19	44	47	31	25	42	178	
5	19	44	44	29	24	42	116	
6	19	44	45	29	24	42	129	
8	19	42	46	29	25	43	46	
9	19	43	46	29	24	43	90	
10	19	42	46	29	24	43	43	
11	19	44	47	29	24	41	166	
16	20	43	47	29	24	43	222	
17	19	43	46	29	24	43	90	
18	19	42	46	29	24	43	43	
19	19	42	45	29	24	42	32	
20	19	42	45	28	24	42	29	
21	19	44	44	29	24	42	116	
22	19	44	44	29	25	42	119	
23	19	42	46	29	24	43	43	
24	19	42	46	29	24	43	43	
25	19	42	46	29	25	43	46	
26	19	42	46	29	25	44	47	

Obrázek 5.6 MS Access databáze

5.3 Program pro vyhledávání

Podoba grafického uživatelského rozhraní

Druhý program umožňuje vyhledávání v MS Access databázi. Po spuštění programu uživatel zvolí rychlé nebo několikanásobné vyhledávání. Pokud je zvoleno rychlé vyhledávání, uživatel vybere požadovanou bakterii (obrázek 5.6), zadá počty G-C párů u konkrétních alel (obrázek 5.7) a program pak zpětně vyhledává konkrétní melt typ a příslušné sekvenční typy. Tyto výsledky může uživatel pouze zobrazit v programu nebo uložit v podobě Excel tabulky. Několikanásobné vyhledávání bylo do programu doplněno především za účelem zrychlení práce v laboratorním provozu. Uživatel může místo zdoluhavého klikání nahrát Excel soubor s přesně definovanou strukturou (obrázek 5.8), kde poslední sloupce uvádí jak název bakterie, tak všech 6 nebo 8 alel potřebných pro vyhledání melt a sekvenčních typů. Výsledky lze poté rovnou uložit do databáze vyšetření pacientů, kde budou sloužit k přesnější identifikaci patogenní bakterie a rychlému a levnému stanovení diagnózy pacienta.

Program načítá z databáze data v datovém typu cell array (česky buňková pole) a poté převádí buňky obsahující čísla na číselný datový typ double a buňky obsahující text (zpravidla hlavičku a sloupec pro poznámky) na strings (česky textový řetězec). Převod na číselný datový typ je potřeba provést, aby bylo možno s čísly dále pracovat. Hledané melt a sekvenční typy jsou získány také v datovém typu double. Musí však být doplněny textové popisky a rozdílné datové typy opět znemožňují ukládání výsledků. Opět tedy musí být výstup převeden na společný datový typ cell array (česky buňková pole).

Vyhledávání MELT a ST

☒ **Rychlé vyhledávání**

ESCO *Escherichia coli*

adk *KLPN Klebsiella pneumoniae* A263 recA91

0 0 0 0 0 0 0 0

☐ **Několikanásobné vyhledávání**

☐

Obrázek 5.7 Uživatelské rozhraní programu pro vyhledávání

Vyhledávání MELT a ST

☒ **Rychlé vyhledávání**

ESCO *Escherichia coli*

adk322 fumC327 gyrB174 icd174 purA263 recA91

45 62 26 0 47 0 0 0

0
43
45
46
47

Obrázek 5.8 Nabídka možných kombinací alel

B	C	D	E	F	G	H	I	J	K
Datum odběru	Klinika	lokalizace vysvětlení	Bakterie	infB 72	mdh 11	phoE 20	rpoB 22	tonB 26	tonB 28
4.12.2017	DIK	rektum	KLPN	13	9	19	45	41	55
9.12.2017	KIGOPL	mož	KLPN	12	9	20	45	41	55
12.12.2017	KIGOPL	rektum	KLPN	12	9	20	45	41	55
19.12.2017	KIGOPL	rektum	KLPN	13	9	20	44	40	56
17.12.2017	IHOK	hemokultura	KLPN	13	9	19	45	41	55
19.12.2017	IHOK	mož	KLPN	12	9	20	45	41	55
19.12.2017	KIGOPL	hemokultura	KLPN	12	9	20	44	40	55
20.12.2017	KIGOPL	rektum	KLPN	13	9	20	44	40	56
22.12.2017	KIGOPL	mož	KLPN	13	9	20	44	40	56

Obrázek 5.9 Podoba souboru vloženého při několikanásobném vyhledávání

6 Zhodnocení

Testování klíče v laboratorním provozu a navržení možných vylepšení bylo provedeno již v průběhu práce na prvním programu. Vylepšení, která byla aplikována do programu jsou popsána v kapitole 5. Celý nástroj byl také průběžně konzultován, aby odpovídal co nejvíce požadavkům uživatelů. Druhý program nástroje vyhledávající melt a sekvenční typy umožňuje přístup k datům, která lze dále vyhodnocovat.

Statistické zpracování dat

V květnu 2018 čítá databáze sekvenční typy celkem pěti bakterií. Konkrétně se jedná o *Enterococcus faecalis* (zkráceně ENFA), *Enterococcus faecium* (ENFI), *Escherichia coli* (ESCO), *Klebsiella pneumoniae* (KLPN) a *Staphylococcus aureus* (STAU).

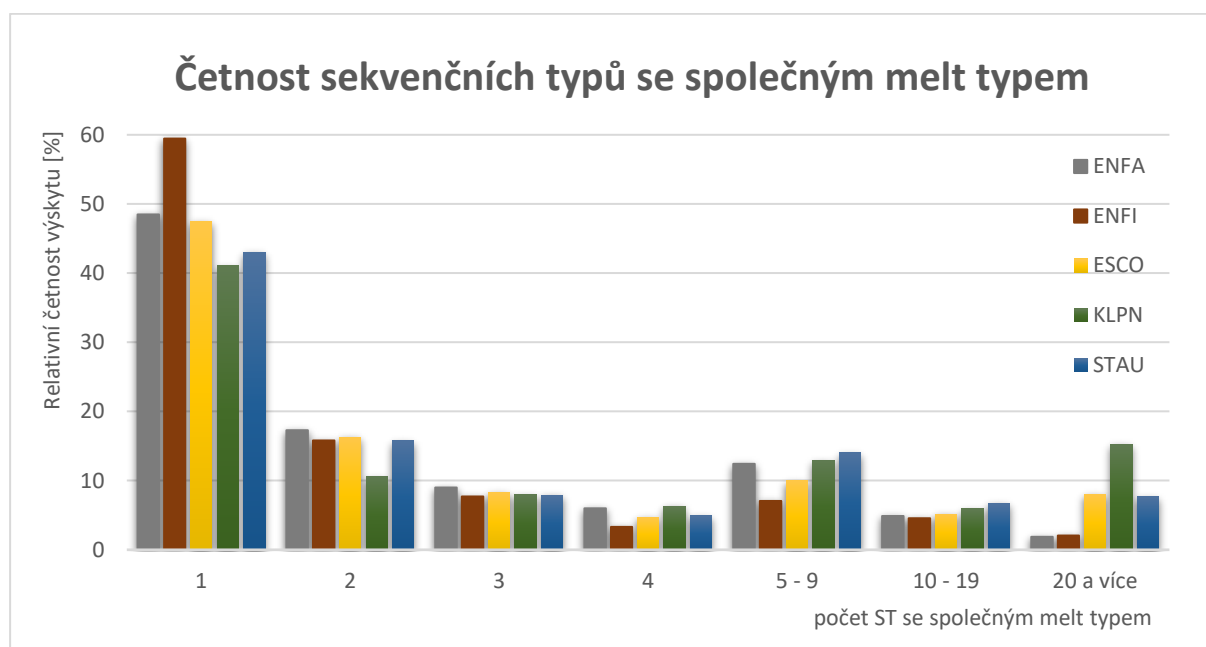
Pro každou bakterii lze zpětně dohledat kolik sekvenčních typů přísluší konkrétnímu melt typu. Na základě těchto údajů lze určit efektivnost metody nebo bakterie porovnat a určit jejich podobnost. Celkové počty sekvenčních a melt typů pro jednotlivé bakterie jsou uvedeny v tabulce 1. Pro všechny melt typy u všech pěti bakterií byly spočteny příslušné sekvenční typy a vyjádřeny jako absolutní a relativní četnosti sekvenčních typů se společným melt typem. Kompletní tabulka je uvedena v elektronické příloze, tabulka 2 ji přehledně shrnuje. Pro názornější zobrazení byl z tabulky 2 vytvořen graf relativních četností sekvenčních typů, který je zobrazen na obrázku 6.1. Z tabulky i z grafu je možné vyčíst, že i když se absolutní četnosti bakterií výrazně liší, relativní četnosti zastoupení sekvenčních typů pro jeden melt typ mají velmi podobné. U všech bakterií převládá poměr jediného sekvenčního typu na jeden melt typ se zastoupením 40-60 %. Dva sekvenční typy připadnou na jeden melt typ u zhruba 10-17 % případů. Tyto výsledky jsou důkazem, že nová metoda je opravdu efektivní. Místo sekvenace DNA za tisíce, byla provedena PCR za koruny a ve výsledku můžeme až u 60 % melt typů zjistit přesný sekvenční typ a u 20 % jej omezit na pouhé dva sekvenční typy. Z grafu také vyplývá, že bakterie *Enterococcus faecalis* (ENFA) a *Enterococcus faecium* (ENFI) jsou pro tuto metodu velmi vhodné, neboť mají nejvyšší procentuální zastoupení jediného sekvenčního typu na jeden melt typ a pouze dvěma procentům melt typů přísluší více než 20 sekvenčních typů. Naopak bakterie *Klebsiella pneumoniae* (KLPN) je nejméně vhodná, neboť u ní je procentuální zastoupení jediného sekvenčního typu na jeden melt typ nejnižší a zastoupení u více než dvaceti sekvenčních typů nejvyšší.

Tabulka 1 Počty sekvenčních a melt typů pro jednotlivé bakterie

	ENFA	ENFI	ESCO	KLPN	STAU
počet sekvenčních typů	854	1420	7496	3228	4609
počet melt typů	266	481	871	302	507

Tabulka 2 Absolutní (m_i) a relativní (r_i) četnost daného počtu ST se společným melt typem

počet ST/melt	ENFA		ENFI		ESCO		KLPN		STAU	
	m_i	r_i [%]	m_i	r_i [%]	m_i	r_i [%]	m_i	r_i [%]	m_i	r_i [%]
1	129	48,5	286	59,5	414	47,5	124	41,1	218	43,0
2	46	17,3	76	15,8	142	16,3	32	10,6	80	15,8
3	24	9,0	37	7,7	72	8,3	24	7,9	40	7,9
4	16	6,0	16	3,3	41	4,7	19	6,3	25	4,9
5 - 9	33	12,4	34	7,1	87	10,0	39	12,9	71	14,0
10 - 19	13	4,9	22	4,6	45	5,2	18	6,0	34	6,7
20 a více	5	1,9	10	2,1	70	8,0	46	15,2	39	7,7



Obrázek 6.1 Graf četností sekvenčních typů se společným melt typem

Pro určení podobnosti bakterií z hlediska zastoupení sekvenčních a melt typů, byla následně vypočtena korelace mezi jednotlivými bakteriemi. Pro každé dvojice bakterií byly spočteny Spearmanovy korelační koeficienty (tabulka 2). Z tabulky 2 vyplývá, že nejpodobnější bakterie z hlediska zastoupení sekvenčních a melt typů jsou *Enterococcus faecium* a *Escherichia coli*. Druhé nejpodobnější jsou si *Enterococcus faecalis* a *Staphylococcus aureus*. Z tabulky 2 také vyplývá, že naopak nejméně podobné jsou si bakterie *Staphylococcus aureus* a *Escherichia coli*, následně *Staphylococcus aureus* a *Klebsiella pneumoniae* a na třetím místě *Escherichia coli* a *Klebsiella pneumoniae*.

Tabulka 3 Spearmanovy korelační koeficienty

	ENFA	ENFI	ESCO	KLPN	STAU
ENFA	1,000				
ENFI	0,613	1,000			
ESCO	0,608	0,641	1,000		
KLPN	0,610	0,618	0,511	1,000	
STAU	0,625	0,619	0,370	0,448	1,000

7 Závěr

Cílem práce bylo provedení literární rešerše o základních metodách pro typizaci bakterií se zaměřením na metody Multi Locus Sequence Typing a Minim Typing, včetně diskuse laboratorních technik, jež metody typizace využívají. Dále návrh základního výpočetního klíče pro převod sekvenčního typu na příslušející melt typ, jeho implementace v prostředí Matlab, rozšíření nástroje o možnost vytvářet a spravovat databáze melt typů, vytvořit přívětivé rozhraní a otestovat jej v laboratorním provozu.

Práce v úvodu popisuje jak základní molekulárně-biologické metody (elektroforetické techniky, polymerázovou řetězovou reakci, sekvenaci DNA), tak i metody bakteriální typizace, které z nich vychází. U každé metody jsou diskutovány výhody a nevýhody. U drtivé většiny z nich je však hlavním problémem finanční nebo časová náročnost. Za tímto účelem byla navržena metoda typizace umožňující srovnání sekvenčních a PCR dat, kterou se zabývá tato práce. Nástroj spravující databázi melt typů umožňuje vyhledávání konkrétních melt typů a příslušných sekvenčních typů, které následně slouží k přesné identifikaci patogenní bakterie a rychlému a levnému stanovení diagnózy pacienta.

Celý nástroj byl průběžně konzultován, aby odpovídal co nejvíce požadavkům biologů z Centra molekulární biologie a genové terapie Fakultní nemocnice Brno. Tím, že algoritmus na základě načtených dat přiřadí sekvenčním typům příslušný melt typ a umožní uložení klíče do databáze, ušetří biologům spoustu času. Doposud trávili dlouhé hodiny přiřazováním příslušných melt typů ručně. Stále je v praxi je potřeba věnovat čas úpravě souborů, které mají být do programu načteny. Sekvence FASTA souborů by měly být zarovnané, hlavička každé sekvence musí obsahovat jméno genu včetně velkých a malých písmen, podtržítka a číslo alely. MLST databáze musí být exportovaná do excel souboru s jediným listem (případně musí být na prvním listu a další listy musejí být prázdné). Sloupce s geny, které nemají být použity musí být odstraněny stejně jako všechny zbytečné popisky a první sloupec musí definovat sekvenční typ. Soubor s primery musí mít rovněž přesně danou strukturu. Přesto úprava souborů do náležité podoby zabere jen zlomek času oproti ručnímu přiřazování melt typů. Téměř všechny bakterie mají totiž více než 1000 sekvenčních typů a například bakterie *Escherichia coli* téměř 7500.

Testování klíče v laboratorním provozu a navržení možných vylepšení bylo provedeno již v průběhu práce na prvním programu. Vylepšení, která byla aplikována do programu jsou popsána v kapitole 5. Druhý program nástroje vyhledávající melt a sekvenční typy umožňuje přístup k datům, která lze dále vyhodnocovat. Pro každou bakterii lze zpětně dohledat kolik sekvenčních typů přísluší konkrétnímu melt typu. Na základě těchto údajů lze bakterie porovnat a určit jejich podobnost. Pro všechny melt typy u všech pěti bakterií byly spočteny příslušné sekvenční typy a vyjádřeny jako absolutní a relativní četnosti sekvenčních typů se společným

melt typem. Z výsledků bylo možné vyčíslit, že i když absolutní četnosti bakterií se výrazně liší, relativní četnosti zastoupení sekvenčních typů pro jeden melt typ mají velmi podobné. U všech bakterií převládal poměr jediného sekvenčního typu na jeden melt typ a melt typy které odpovídají více než 5 sekvenčním typům tvoří pouze malé procento případů. Je tak možné levně a přesně identifikovat.

.

Seznam použité literatury

- [1] LEFFLER, D.A. a T.J. LAMONT. Clostridium difficile Infection. *New England Journal of Medicine*. 2015, **372**(16), 1539-1548. DOI: 10.1056/NEJMoa1408913.
- [2] LAXMINARAYAN, R., A. DUSE, Ch. WATTAL et al. Antibiotic resistance-the need for global solutions. *The Lancet Infectious Diseases*. 2013, **13**(12), 1057-1098. DOI: 10.1016/S1473-3099(13)70318-9.
- [3] BEJ, A.K., M.H. MAHBUBANI a R.M. ATLAS. Amplification of Nucleic Acids by Polymerase Chain Reaction (PCR) and Other Methods and their Applications. *Critical Reviews in Biochemistry and Molecular Biology*. 1991, **26**(3-4), 301-334. DOI: 10.3109/10409239109114071.
- [4] ALBERTS, B. *Základy buněčné biologie: úvod do molekulární biologie buňky*. 2. vyd. Ústí nad Labem: Espero, 2004. ISBN 80-902906-2-0.
- [5] MAXAM, A.M. a W. GILBERT. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 1977, **74**(2), 560-4. DOI: 10.1073/pnas.74.2.560.
- [6] SANGER, F, S. NICKLEN a A.R. COULSON. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977, **74**(12), 5463-5467. DOI: 10.1073/pnas.74.12.5463.
- [7] ZHOU, X., L. REN, Q. MENG, Y. LI, Y. YU a J. YU. The next-generation sequencing technology and application. *Protein & Cell*. Higher Education Press, 2010, **1**(6), 520-536. DOI: 10.1007/s13238-010-0065-3.
- [8] XUAN, J., Y. YU, T. QING, L. GUO a L. SHI. Next-generation sequencing in the clinic: Promises and challenges: Promises and challenges. *Cancer Letters*. Elsevier, 2013, **340**(2), 284-295. DOI: 10.1016/j.canlet.2012.11.025.
- [9] METZKER, M.L. Sequencing technologies — the next generation. *Nature Reviews Genetics*. Nature Publishing Group, 2010, **11**(1), 31-46. DOI: 10.1038/nrg2626.
- [10] Different Platforms of sequencing, Part IV. *Playing with Sequences* [online]. Mountain View: Google, 2012 [cit. 2018-05-15]. Dostupné z: <http://genomesequense.blogspot.cz/>
- [11] LIU, L., Y. LI, S. LI et al. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*. Hindawi Publishing Corporation, 2012, **2012**(11), 1-11. DOI: 10.1155/2012/251364.
- [12] ANSORGE, W.J. Next-generation DNA sequencing techniques. *New Biotechnology*. 2009, **25**(4), 195-203. DOI: 10.1016/j.nbt.2008.12.009.

- [13] HARRIS, T.D., P.R. BUZBY, H. BABCOCK et al. Single-Molecule DNA Sequencing of a Viral Genome. *Science*. 2008, **320**(5872), 106-109. DOI: 10.1126/science.1150427.
- [14] TUCKER, T., M. MARRA a J.M. FRIEDMAN. Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. *The American Journal of Human Genetics*. 2009, **85**(2), 142-154. DOI: 10.1016/j.ajhg.2009.06.022.
- [15] MCCARTHY, A. Third Generation DNA Sequencing: Pacific Biosciences' Single Molecule Real Time Technology. *Chemistry & Biology*. 2010, **17**(7), 675-676. DOI: 10.1016/j.chembiol.2010.07.004.
- [16] GOTTWALDOVÁ, J. *Elektroforetické techniky* [online]. OKB FN Brno, 2013. Dostupné také z: <https://is.muni.cz/el/1411/jaro2013/BLIT0222p/um>. Studijní materiály. Masarykova univerzita.
- [17] ŠPANOVÁ, A., B. RITTICH, M.J. BENEŠ a D. HORÁK. Ferrite supports for isolation of DNA from complex samples and polymerase chain reaction amplification. *Journal of Chromatography A*. 2005, **1080**(1), 93-98. DOI: 10.1016/j.chroma.2005.05.006.
- [18] MALORNY, B., Ch. LÖFSTRÖM, M. WAGNER, N. KRÄMER a J. HOORFAR. Enumeration of Salmonella bacteria in food and feed samples by real-time PCR for quantitative microbial risk assessment. *Applied and Environmental Microbiology*. 2008, **74**(5), 1299-1304. DOI: 10.1128/AEM.02489-07.
- [19] POSTOLLEC, F., H. FALENTIN, S. PAVAN, J. COMBRISSEON a D. SOHIER. Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiology*. 2011, **28**(5), 848-861. DOI: 10.1016/j.fm.2011.02.008.
- [20] MASCO, L., T. VANHOUTTE, R. TEMMERMAN, J. SWINGS a G. HUYS. Evaluation of real-time PCR targeting the 16S rRNA and recA genes for the enumeration of bifidobacteria in probiotic products. *International Journal of Food Microbiology*. 2007, **113**(3), 351-357. DOI: 10.1016/j.ijfoodmicro.2006.07.021.
- [21] TONG, S.Y.C a P.M. GIFFARD. Microbiological Applications of High-Resolution Melting Analysis. *Journal of Clinical Microbiology*. 2012, **50**(11), 3418-3421. DOI: 10.1128/JCM.01709-12.
- [22] High-Resolution Melting. *The Wittwer Lab for DNA Analysis* [online]. Salt Lake City: Wittwer Lab, ©2006-2018 [cit. 2018-05-15]. Dostupné z: https://www.dna.utah.edu/Hi-Res/TOP_Hi-Res%20Melting.html
- [23] DELMEE, M., M. HOMEL a G. WAUTERS. Serogrouping of Clostridium difficile strains by slide agglutination. *Journal of Clinical Microbiology*. American Society for Microbiology (ASM), 1985, **21**(3), 323-327. ISSN 0095-1137.
- [24] WÜST, J., N.M. SULLIVAN, U. HARDEGGER a T.D. WILKINS. Investigation of an outbreak of antibiotic-associated colitis by various typing methods. *Journal of clinical microbiology*. 1982, **16**(6), 1096-1101. ISSN 0095-1137.

- [25] VAN BELKUM, A., P.T. TASSIOS, L. DIJKSHOORN et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology and Infection*. 2007, **13**(3), 1-46. DOI: 10.1111/j.1469-0691.2007.01786.x.
- [26] WALL, R., G. FITZGERALD, S. HUSSEY, T. RYAN, B. MURPHY, P. ROSS a C. STANTON. Genomic diversity of cultivable *Lactobacillus* populations residing in the neonatal and adult gastrointestinal tract. *FEMS microbiology ecology*. 2007, **59**(1), 127-137. DOI: 10.1111/j.1574-6941.2006.00202.x.
- [27] CALABI, E., S. WARD, B. WREN et al. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Molecular Microbiology*. 2001, **40**(5), 1187-1199. DOI: 10.1046/j.1365-2958.2001.02461.x.
- [28] SABAT, A.J., A. BUDIMIR, D. NASHEV, R. SÁ-LEÃO, J.M. VAN DIJL, F. LAURENT, H. GRUNDMANN a A.W. FRIEDRICH. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance*. European Centre for Disease Prevention and Control, 2013, **18**(4), 17-30.
- [29] BRHELOVA, E., I. KOCMANOVA, Z. RACIL, M. HANSLIANOVA, M. ANTONOVA, J. MAYER a M. LENGEROVA. Validation of Minim typing for fast and accurate discrimination of extended-spectrum, beta-lactamase-producing *Klebsiella pneumoniae* isolates in tertiary care hospital. *Diagnostic Microbiology and Infectious Disease*. 2016, **86**(1), 44-49. DOI: 10.1016/j.diagmicrobio.2016.03.010.
- [30] KRISTJÁNSSON, M., M.H. SAMORE, D.N. GERDING, P.C. DEGIROLAMI, K.M. BETTIN, A.W. KARCHMER a R.D. ARBEIT. Comparison of restriction endonuclease analysis, ribotyping, and pulsed-field gel electrophoresis for molecular differentiation of *Clostridium difficile* strains. *Journal of clinical microbiology*. 1994, **32**(8), 1963-1969.
- [31] SINGH, S., P. GOSWAMI, R. SINGH, K.J. HELLER a K.J. HELLER. Application of molecular identification tools for *Lactobacillus*, with a focus on discrimination between closely related species: A review. *LWT - Food Science and Technology*. 2009, **42**(2), 448-457. DOI: 10.1016/j.lwt.2008.05.019.
- [32] KEIM, P., L.B. PRICE, A.M. KLEVYTSKA, K.L. SMITH, J.M. SCHUPP, R. OKINAKA, P.J. JACKSON a M.E. HUGH-JONES. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *Journal of Bacteriology*. 2000, **182**(10), 2928-2936. DOI: 10.1128/JB.182.10.2928-2936.2000.
- [33] MUELLER, U.G. a L.L.R WOLFENBARGER. AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*. 1999, **14**(10), 389-394. DOI: 10.1016/S0169-5347(99)01659-6.
- [34] GÜRTLER, V. Typing of *Clostridium difficile* strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. *Journal of general microbiology*. 1993, **139**(12), 3089-3097. DOI: 10.1099/00221287-139-12-3089.

- [35] RUPNIK, M., M.H. WILCOX a D.N. GERDING. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nature Reviews Microbiology*. 2009, **7**(7), 526-536. DOI: 10.1038/nrmicro2164.
- [36] HADRYIS, H., M. BALICK a B. SCHIERWATER. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology*. 1992, **1**(1), 55-63. DOI: 10.1111/j.1365-294X.1992.tb00155.x.
- [37] KATO, H., H. KATO, M. NAKAMURA, Y. IWASHIMA, A. NAKAMURA a R. UEDA. Rapid analysis of Clostridium difficile strains recovered from hospitalized patients by using the slpA sequence typing system. *Journal of Infection and Chemotherapy*. 2009, **15**(3), 199-202. DOI: 10.1007/s10156-009-0681-0.
- [38] MAIDEN, M.C.J., J.A. BYGRAVES, E. FEIL et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*. 1998, **95**(6), 3140-3145. DOI: 10.1073/pnas.95.6.3140.
- [39] LEMEE, L., A. DHALLUIN, M. PESTEL-CARON, J.F. LEMELAND a J.L. PONS. Multilocus Sequence Typing Analysis of Human and Animal Clostridium difficile Isolates of Various Toxigenic Types. *Journal of Clinical Microbiology*. 2004, **42**(6), 2609-2617. DOI: 10.1128/JCM.42.6.2609-2617.2004.
- [40] ANDERSSON, P., S.Y.C. TONG, J.M. BELL, J.D. TURNIDGE, P.M. GIFFARD a I. MOKROUSOV. Minim Typing – A Rapid and Low Cost MLST Based Typing Tool for Klebsiella pneumoniae. *PLoS ONE*. 2012, **7**(3), 33530. DOI: 10.1371/journal.pone.0033530.
- [41] Introduction. *Multilocus sequence typing* [online]. Londýn: Imperial College London, ©2018 [cit. 2018-05-15]. Dostupné z: <http://beta.mlst.net/Instructions/default.html>
- [42] *PubMLST.org* [online]. Oxford: University of Oxford, ©2018 [cit. 2018-05-15]. Dostupné z: <https://pubmlst.org/databases/>
- [43] HANSEN, D.S., R. SKOV, J.V. BENEDÍ, V. SPERLING a H.J. KOLMOS. Klebsiella typing: pulsed-field gel electrophoresis (PFGE) in comparison with O:K-serotyping. *Clinical microbiology and infection*. 2002, **8**(7), 397-404. DOI: 10.1046/j.1469-0691.2002.00411.x.
- [44] CLABOTS, C.R., S. JOHNSON, K.M. BETTIN, P.A. MATHIE, M.E. MULLIGAN, D.R. SCHABERG, L.R. PETERSON a D.N. GERDING. Development of a rapid and efficient restriction endonuclease analysis typing system for Clostridium difficile and correlation with other typing systems. *Journal of Clinical Microbiology*. 1993, **31**(7), 1870-1875. ISSN 0095-1137.
- [45] POWELL, W., M. MORGANTE, Ch. ANDRE, M. HANAFEY, J. VOGEL, S. TINGEY a A. RAFALSKI. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite)

- markers for germplasm analysis. *Molecular Breeding*. 1996, **2**(3), 225-238. DOI: 10.1007/BF00564200.
- [46] MARSH, J.W., M.M. O'LEARY, K.A. SHUTT, A.W. PASCULLE, S. JOHNSON, D.N. GERDING, C.A. MUTO a L.H. HARRISON. Multilocus variable-number tandem-repeat analysis for investigation of *Clostridium difficile* transmission in hospitals. *Journal of Clinical Microbiology*. 2006, **44**(7), 2558-2566. DOI: 10.1128/JCM.02364-05.
- [47] VAN DEN BERG, R.J., I. SCHAAP, K.E. TEMPLETON, C.H.W. KLAASSEN a E.J. KUIJPER. Typing and subtyping of *Clostridium difficile* isolates by using multiple-locus variable-number tandem-repeat analysis. *Journal of Clinical Microbiology*. 2007, **45**(3), 1024-1028. DOI: 10.1128/JCM.02023-06.
- [48] VAN DEN BERG, R.J., E.C.J. CLAAS, D.H. OYIB, C.H.W. KLAASSEN, L. DIJKSHOORN, J.S. BRAZIER a E.J. KUIJPER. Characterization of Toxin A-Negative, Toxin B-Positive *Clostridium difficile* Isolates from Outbreaks in Different Countries by Amplified Fragment Length Polymorphism and PCR Ribotyping. *Journal of Clinical Microbiology*. 2004, **42**(3), 1035-1041. DOI: 10.1128/JCM.42.3.1035-1041.2004.
- [49] INDRA, A., S. HUHULESCU, M. SCHNEEWEIS et al. Characterization of *Clostridium difficile* isolates using capillary gel electrophoresis-based PCR ribotyping. *Journal of Medical Microbiology*. 2008, **57**(11), 1377-1382. DOI: 10.1099/jmm.0.47714-0.
- [50] RUPNIK, M., V. AVESANI, M. JANC, C. VON EICHEL-STREIBER a M. DELMÉE. A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *Journal of clinical microbiology*. 1998, **36**(8), 2240-2247. ISSN 0095-1137.
- [51] LANINI, S., S. D'AREZZO, V. PURO et al. Molecular epidemiology of a *Pseudomonas aeruginosa* hospital outbreak driven by a contaminated disinfectant-soap dispenser. *PLoS ONE*. 2011, **6**(2), 1-10. DOI: 10.1371/journal.pone.0017064.

Seznam použitých zkratk

AFLP	Amplified Fragment Length Polymorphism	polymorfismus délek amplifikovaných fragmentů
bp	base pair	komplementární pár bází
DNA	deoxyribonucleic acid	kyselina deoxyribonukleová
dsDNA	double-stranded DNA	dvouvláknová DNA
ssDNA	single-stranded DNA	jednovláknová DNA
ENFA	Enterococcus faecalis	
ENFI	Enterococcus faecium	
ESCO	Escherichia coli	
FASTA	fast alignment	sada programů umožňující srovnávání sekvencí DNA a proteinů
HRM	High Resolution Melting	analýza s vysokou rozlišovací schopností
KLPN	Klebsiella pneumoniae	
MLST	multilocus sequence typing	typizace dle sekvencí mnoha lokusů
MLVA	multilocus variable number	multilokusová analýza tandemových repetit
NGS	Next-Generation Sequencing	sekvenační metody nové generace
PaLoc	Pathogenicity locus	lokus patogenicity
PFGE	Pulsed Field Gel Electrophoresis	pulzní gelová elektroforéza
REA	Endonuclease Restriction Analysis	restrikční endonukleázová analýza
RFLP	Restriction Fragment Length Polymorphism	polymorfismus délek restrikčních fragmentů
<i>slp</i> AST	Surface Layer Protein A Gene Sequence Typing	typizace dle variability sekvence genu <i>slpA</i>
SMRT	Single Molecule Real-Time Sequencing	sekvenace jednotlivých molekul v reálném čase
SNP	Single Nucleotide Polymorphism	jednonukleotidový polymorfismus
SOLiD	Sequencing by Oligo Ligation and Detection	sekvencování oligonukleotidů na základě ligace
ST	Sequence Type	sekvenční typ

<i>STAU</i>	Staphylococcus aureus	
<i>TcdA</i>	toxin A <i>Clostridium difficile</i>	toxin A <i>Clostridium difficile</i>
<i>TcdB</i>	toxin B <i>Clostridium difficile</i>	toxin B <i>Clostridium difficile</i>
<i>Tm</i>	Melting Temperature	teplota tání
tSMS	True Single Molecule Sequencing	tzv. přesné sekvencování jednotlivých molekul DNA

Seznam elektronických příloh

Program pro správu MS Access databáze

Program pro vyhledávání

Testovací data

Kompletní tabulka absolutních a relativních četností daného počtu ST se společným melt typem